

# How can Fleiss' Kappa be calculated using Excel?

Authored by  
**stats writer**

April 17, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can Fleiss' Kappa be calculated using Excel?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136356>

Fleiss' Kappa is a statistical measure used to assess the agreement between multiple raters on a categorical variable. It is commonly used in fields such as psychology, sociology, and medicine. Excel can be used to calculate Fleiss' Kappa by following these steps:

1. Create a table in Excel, with each row representing a subject and each column representing a rater.
2. Assign numerical codes to the categories being rated (e.g. 1 for "agree", 2 for "disagree", etc.).
3. Enter the ratings for each rater in the corresponding cells of the table.
4. Calculate the total number of raters ( $n$ ) and the total number of categories ( $k$ ).
5. Calculate the overall proportion of agreement ( $P$ ) by summing the number of ratings in each category and dividing by the total number of ratings.
6. Calculate the proportion of agreement that would be expected by chance ( $P_e$ ) by summing the squared proportions for each category and dividing by the number of raters.
7. Calculate Fleiss' Kappa ( $K$ ) using the formula:  $K = (P - P_e) / (1 - P_e)$
8. The resulting value of  $K$  will range from -1 to 1, with 1 indicating perfect agreement and 0 indicating no agreement beyond chance.

## Calculate Fleiss' Kappa in Excel

**Fleiss' Kappa is a way to measure the degree of agreement between three or more raters when the raters are assigning categorical ratings to a set of items.**

**Fleiss' Kappa ranges from 0 to 1 where:**

**0 indicates no agreement at all among the raters. 1 indicates perfect inter-rater agreement.**

**This tutorial provides an example of how to calculate Fleiss' Kappa in Excel.**

### Example: Fleiss' Kappa in Excel

Suppose 14 individuals rate 10 different products on a scale of Poor to Excellent.

The following screenshot displays the total ratings that each product received:

	A	B	C	D	E	F	G	H	I
1			<b>Rating</b>						
2		<b>Product ID</b>	Poor	Weak	Average	Good	Excellent		
3		1	0	0	0	0	14		
4		2	0	2	6	4	2		
5		3	0	0	3	5	6		
6		4	0	3	9	2	0		
7		5	2	2	8	1	1		
8		6	7	7	0	0	0		
9		7	3	2	6	3	0		
10		8	2	5	3	2	2		
11		9	6	5	2	1	0		
12		10	0	2	2	3	7		
13									
14									
15									
16									
17									
18									
19									
20									
21									

The following screenshot shows how to calculate Fleiss' Kappa for this data in Excel:

J3     $= (C3^2 - C3) + (D3^2 - D3) + (E3^2 - E3) + (F3^2 - F3) + (G3^2 - G3)$

	A	B	C	D	E	F	G	H	I	J	K	L	
1			Rating										
2		Product ID	Poor	Weak	Average	Good	Excellent		Sum	sum sq. c	sum sq. c / (14*13)		
3		1	0	0	0	0	14		14	182	1.0000		
4		2	0	2	6	4	2		14	46	0.2527		
5		3	0	0	3	5	6		14	56	0.3077		
6		4	0	3	9	2	0		14	80	0.4396		
7		5	2	2	8	1	1		14	60	0.3297		
8		6	7	7	0	0	0		14	84	0.4615		
9		7	3	2	6	3	0		14	44	0.2418		
10		8	2	5	3	2	2		14	32	0.1758		
11		9	6	5	2	1	0		14	52	0.2857		
12		10	0	2	2	3	7		14	52	0.2857		
13													
14		sum	20	28	39	21	32				3.7802	sum	
15		sum/140	0.143	0.200	0.279	0.150	0.229				0.37802	sum/10	
16		(sum/140)^2	0.0204	0.0400	0.0776	0.0225	0.0522	0.2128					
17								Sum					
18		Fleiss' K	0.2099										
19													
20													
21													

The trickiest calculations in this screenshot are in column J. The formula used for these calculations is shown in the text box near the top of the screen.

Note that the Fleiss' Kappa in this example turns out to be 0.2099. The actual formula used to calculate this value in cell C18 is:

$$\text{Fleiss' Kappa} = (0.37802 - 0.2128) / (1 - 0.2128) = 0.2099.$$

Although there is no formal way to interpret Fleiss' Kappa, the following values show how to interpret Cohen's Kappa, which is used to assess the level of inter-rater agreement between just two raters:

**< 0.20 | Poor.21 - .40 | Fair.41 - .60 | Moderate.61 - .80 |  
Good.81 - 1 | Very Good**

**Based on these values, Fleiss' Kappa of 0.2099 in our example would be interpreted as a "fair" level of inter-rater agreement.**

ARABPSYCHOLOGY.COM