

How can I calculate cosine similarity in Excel?

Authored by
stats writer

May 4, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate cosine similarity in Excel?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=142628>

Cosine similarity is a mathematical measure used to determine the similarity between two vectors in a multi-dimensional space. In Excel, this can be calculated using the COS function, which computes the cosine of an angle given in radians. To calculate the cosine similarity between two vectors, the user must first convert the data into a numerical format and then use the COS function to calculate the cosine of the angle between the two vectors. This will result in a value between -1 and 1, where a higher value indicates a higher degree of similarity between the two vectors. By using this method, Excel can be utilized as a convenient tool for calculating cosine similarity, making it a useful tool for various data analysis and comparison tasks.

Calculate Cosine Similarity in Excel

Cosine Similarity is a measure of the similarity between two vectors of an inner product space.

For two vectors, A and B, the Cosine Similarity is calculated as:

$$\text{Cosine Similarity} = \frac{\sum A_i B_i}{(\sqrt{\sum A_i^2} \sqrt{\sum B_i^2})}$$

This tutorial explains how to calculate the Cosine Similarity between vectors in Excel.

Cosine Similarity Between Two Vectors in Excel

Suppose we have the following two vectors in Excel:

	A	B	C	D	E	F	G
1	Dataset A	Dataset B					
2	23	17					
3	34	18					
4	44	22					
5	45	26					
6	42	26					
7	27	29					
8	33	31					
9	34	30					
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							

We can use the following formula to calculate the Cosine Similarity between the two vectors in Excel:

=SUMPRODUCT(A\$2:A\$9,B2:B9)/(SQRT(SUMSQ(B2:B9))*SQRT(SUMSQ(\$A\$2:\$A\$9)))

The following screenshot shows how to use this formula:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Dataset A	Dataset B		Cosine Similarity	0.965195	=SUMPRODUCT(A\$2:A\$9,B2:B9)/(SQRT(SUMSQ(B2:B9))*SQRT(SUMSQ(\$A\$2:\$A\$9)))							
2	23	17											
3	34	18											
4	44	22											
5	45	26											
6	42	26											
7	27	29											
8	33	31											
9	34	30											
10													
11													
12													
13													
14													
15													
16													
17													
18													
19													
20													
21													

The Cosine Similarity between the two vectors turns out to be 0.965195.

Note that we interpret the value for Cosine Similarity as follows:

A value of -1 indicates maximum dissimilarity
 A value of 0 indicates that the two vectors are orthogonal
 A value of 1 indicates maximum similarity

A value of 0.965195 indicates a high level of similarity between the two vectors in our example.

The following tutorials explain how to calculate Cosine Similarity in other statistical software:

For an in-depth explanation of Cosine Similarity, refer to the following Wikipedia article:

ARABPSYCHOLOGY.COM