

How can I calculate the Hamming Distance using Excel?

Authored by
stats writer

April 23, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate the Hamming Distance using Excel?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=138474>

The Hamming Distance is a measure of the number of positions in which two strings of equal length differ. It is often used in various fields such as coding theory, computer science, and genetics. Excel, being a powerful spreadsheet program, can be utilized to calculate the Hamming Distance between two strings of binary data. This can be achieved by using the built-in function "BITXOR" to compare the bits of the two strings and then summing up the resulting values. This process can be repeated for multiple pairs of strings to calculate the Hamming Distance for a set of data. By following this method, Excel can be a useful tool for computing the Hamming Distance in a quick and efficient manner.

Calculate Hamming Distance in Excel

The Hamming distance between two vectors is simply the sum of corresponding elements that differ between the vectors.

For example, suppose we have the following two vectors:

x =

y =

The Hamming distance between the two vectors would be 2, since this is the total number of corresponding elements that have different values.

To calculate the Hamming distance between two columns in Excel, we can use the following syntax:

=COUNT(RANGE1)-SUMPRODUCT(--(RANGE1 = RANGE2))

Here's what the formula does in a nutshell:

COUNT finds the total number of observations in the first column. **RANGE1 = RANGE2** compares each pairwise observations between the columns and returns a **TRUE** or **FALSE**. **--** converts **TRUE** and **FALSE** values to 0 and 1. **SUMPRODUCT** finds the sum of all 1's.

This tutorial provides several examples of how to use this calculation in practice.

Example 1: Hamming Distance Between Binary Vectors

The following code shows how to calculate the Hamming distance between two columns in Excel that each contain only two possible values: