

Learn How to Calculate Manhattan Distance Using Excel

Authored by
Mohammed looti

November 3, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Learn How to Calculate Manhattan Distance Using Excel*.
PSYCHOLOGICAL STATISTICS. Retrieved from
<https://statistics.arabpsychology.com/?p=8943>

Introducing the Manhattan Distance: Definition and Context

The [Manhattan distance](#), often formally designated as the **L1 norm** or colloquially as **taxicab geometry**, represents a crucial metric in analytical geometry and data science. Unlike the standard, straight-line distance, which is known as the [Euclidean distance](#), the Manhattan distance strictly measures the distance between two points by summing the lengths of the projections of the segment onto the coordinate axes. This constraint simulates movement only along right angles, mirroring how a taxi must navigate the grid system of a city like Manhattan, New York.

Mathematically, this metric calculates the dissimilarity between two [vectors](#), A and B , by finding the sum of the absolute differences of their Cartesian coordinates across all dimensions. This definition yields a robust and easily interpretable measure of separation, especially valuable in scenarios where orthogonal movement constraints are inherent to the problem or when feature independence is a key assumption in the underlying data structure.

The core mathematical expression is fundamental to implementing this calculation in spreadsheet applications such as [Excel](#). The formula requires iterating through all dimensions, taking the absolute value of the difference between corresponding elements (or coordinates) of the two vectors, and finally aggregating those values. The standard notation succinctly captures this process:

$$\sum |A_i - B_i|$$

where the index i references the i th element in each respective vector, ensuring that every feature contributes positively to the total distance calculated.

Why the L1 Norm Matters: Applications in Data Science

The utility of the Manhattan distance extends significantly beyond theoretical geometry, establishing it as a fundamental tool in various computational disciplines, including advanced statistical modeling and [machine learning algorithms](#). Its preference often emerges in contexts involving high-dimensional data spaces or specialized feature selection tasks, particularly where the presence of outliers could unduly influence metrics that rely on squaring differences, such as the Euclidean or L2 norm.

Analysts frequently choose the Manhattan distance when the input variables exhibit low correlation or when they specifically desire a metric that is less sensitive to extreme deviations. By utilizing the **absolute value** of differences rather than the squared differences, the L1 norm effectively penalizes large errors linearly, contrasting with the quadratic penalty applied by the L2 norm. This characteristic makes it highly suitable for applications in image processing, certain forms of pattern recognition, and key clustering methods, notably the K-Nearest Neighbors (KNN) algorithm.

A critical advantage of the Manhattan distance is its direct interpretability. Since the calculation involves summing the absolute differences, the resulting distance value is expressed in the original units of the features being measured. This transparency offers a significant benefit in analytical contexts where communicating the magnitude of separation in real-world terms is essential for stakeholders or for auditing the underlying data model.

Structuring Your Data for Manhattan Distance in Excel

To practically execute the calculation of the Manhattan distance, we must first properly structure the input data within Microsoft [Excel](#). The process requires establishing the two data points, Vector A and Vector B, as distinct columns within the spreadsheet environment. This vertical layout is optimal as it facilitates the efficient application of Excel's powerful built-in mathematical functions across corresponding elements.

For illustrative purposes, consider an example involving two [vectors](#), A and B, each defined by four distinct numerical features. These features could represent diverse datasets, ranging from quarterly sales figures and physiological measurements to specific experimental outcomes or pixel intensities in a digital image. The setup below demonstrates the necessary organization within the Excel worksheet, where Vector A resides in Column B and Vector B occupies Column C.

	A	B	C	D	E	F	G	H
1	A	B						
2	2	3						
3	6	5						
4	7	5						
5	7	6						
6	8	7						
7	12	7						
8	13	9						
9	16	10						
10	17	11						
11	17	13						
12	18	13						
13	19	15						
14	20	18						
15	22	18						
16	24	19						
17								
18								
19								
20								
21								
22								
23								
24								

Before proceeding to the final aggregation step, the primary intermediate requirement is to compute the **absolute difference** between every corresponding element pair (A_i and B_i). This intermediate step is essential because the definition of the [Manhattan distance](#) mandates that only the magnitude of the difference contributes to the final total, irrespective of which vector component holds the greater value.

Step 1: Calculating the Absolute Differences using ABS()

The calculation of the Manhattan distance hinges entirely on accurately determining the absolute difference for each pair of vector elements. Excel provides an extremely efficient way to manage this requirement through the use of its proprietary function, **ABS()**. This function is designed specifically to take any single numerical input--be it a constant, a formula result, or a cell reference--and invariably return its positive (absolute) value, thereby ensuring non-negativity as required by the L1 norm definition.

We dedicate a new column, conventionally titled "Difference ($|A-B|$)," to systematically store these intermediate absolute values. For the initial row of data (representing $i=1$), the formula entered into

this dedicated column must reference the corresponding cells in Vector A and Vector B. Assuming the data begins in Row 2, the structure of the formula is straightforward: `=ABS(B2-C2)`. This instructs Excel to calculate the difference between the elements and subsequently apply the absolute value rule.

A significant efficiency gain is achieved by utilizing Excel's fill handle. Once the formula is correctly established for the first pair of elements, it can be seamlessly dragged down the column. This action automatically adjusts the cell references for all subsequent rows, completing the calculation of the absolute difference for every pair of components in the dataset. This step prepares the necessary figures for the final summation required to obtain the total [Manhattan distance](#).

	A	B	C	D	E	F	G
1	A	B	Abs(A-B)				
2	2	3	1	<code>=ABS(A2-B2)</code>			
3	6	5	1				
4	7	5	2				
5	7	6	1				
6	8	7	1				
7	12	7	5				
8	13	9	4				
9	16	10	6				
10	17	11	6				
11	17	13	4				
12	18	13	5				
13	19	15	4				
14	20	18	2				
15	22	18	4				
16	24	19	5				
17							
18							
19							
20							
21							
22							
23							
24							

Step 2: Aggregating the Results with the SUM Function

Once all the individual absolute differences have been meticulously calculated and stored, the final step in deriving the Manhattan distance is to perform the aggregation. This critical action satisfies the summation (Σ) component inherent in the fundamental mathematical formula, effectively bringing together the contribution of every dimension into a single, comprehensive metric.

Within [Excel](#), the built-in **SUM()** function is the most direct and efficient mechanism for achieving this total aggregation. By simply selecting the entire range of cells that constitute the "Difference ($|A-B|$)" column, the function computes the cumulative distance traversed along all axes. This grand total precisely quantifies the total dissimilarity, or L1 norm, between the initial Vector A and Vector B across all the features included in the analysis.

For instance, if the intermediate absolute differences are stored in column D, spanning from row 2 through row 5, the definitive formula is entered into a designated output cell (e.g., D6 or D7) as `=SUM(D2:D5)`. The resulting output is a solitary numerical value, such as **51** in our example, which serves as the robust quantification of the spatial or feature separation based on the taxicab geometry.

	A	B	C	D	E	F	G
1	A	B	Abs(A-B)				
2	2	3	1				
3	6	5	1				
4	7	5	2				
5	7	6	1				
6	8	7	1				
7	12	7	5				
8	13	9	4				
9	16	10	6				
10	17	11	6				
11	17	13	4				
12	18	13	5				
13	19	15	4				
14	20	18	2				
15	22	18	4				
16	24	19	5				
17		Sum	51	<code>=SUM(C2:C16)</code>			
18							
19							
20							
21							
22							
23							
24							

As demonstrated by the final calculation, the Manhattan distance between the two input [vectors](#) is determined to be 51. This value represents the total path length when movement is restricted to the axes, providing a clear metric for dissimilarity.

Streamlining Calculations with Advanced Array Formulas

While the step-by-step, three-column methodology (A, B, and |A-B|) offers unparalleled transparency and is ideal for educational purposes or for auditing complex datasets, power users of Excel can leverage the efficiency of a single-cell **array formula** to achieve the same result. An array formula possesses the distinct capability to perform complex, multiple-item calculations across one or more data ranges (arrays) and condense the output into a single result, significantly reducing spreadsheet clutter.

The consolidated array formula for the Manhattan distance cleverly integrates both the **SUM()** and the [ABS\(\) function](#). Assuming that the input ranges for Vector A are B2:B5 and for Vector B are C2:C5, the formula is structured as follows:

```
=SUM(ABS(B2:B5 - C2:C5))
```

A crucial consideration for implementing this method relates to the specific version of Excel being utilized. Historically, such formulas required confirmation by pressing the key combination **Ctrl + Shift + Enter (CSE)** to signal Excel that an array operation was intended. However, modern iterations of Microsoft 365 often incorporate dynamic array handling, automatically recognizing and processing array operations without the mandatory CSE input, which greatly enhances workflow efficiency when dealing with extensive data sets.

The primary benefit of adopting the array method is the substantial improvement in efficiency and the minimization of auxiliary columns within the worksheet. Nonetheless, the multi-step approach remains superior when the core objectives involve detailed auditing of intermediate steps, pedagogical demonstration, or initial instruction in statistical methods.

Beyond the Taxicab: Other Essential Distance Metrics

The [Manhattan distance](#) serves as a foundational metric, yet it is merely one tool in the analytical toolkit used to quantify feature separation between data points. In statistics, robotics, network analysis, and data science, the choice of metric is highly dependent on the nature of the data and the specific objectives of the application. A strong analytical foundation requires familiarity with these diverse alternatives.

Understanding the distinctions between metrics--such as the difference between the L1 norm and the L2 norm--is crucial for robust data modeling and effective interpretation of results. The choice impacts how errors are penalized and how the resulting distance relates to the real-world properties of the system being modeled.

Several other widely employed distance metrics are:

Euclidean Distance: Defined as the straight-line distance, it is the most common metric and is also known as the **L2 norm**.

Chebyshev Distance: This metric calculates the maximum absolute difference across any single coordinate dimension, often referred to as the **L ∞ norm**.

Minkowski Distance: Representing a generalized metric that can encompass both the [Euclidean distance](#) and the Manhattan distance, depending on the chosen parameter (p).

Hamming Distance: Primarily utilized for comparing binary or categorical vectors, measuring the number of positions at which the corresponding symbols are different.

For those interested in expanding their knowledge of geometric and statistical separation measures, further tutorials detail the implementation of these alternative distance calculations directly within Excel: