

How can the dist function be used in R, and what are some examples of its applications?

Authored by
stats writer

May 1, 2024

RECOMMENDED CITATION

stats writer (2024). *How can the dist function be used in R, and what are some examples of its applications?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=141706>

The `dist` function in R is used to calculate the distance between two sets of data points. It can be used for various purposes such as clustering, classification, and data visualization. With the ability to compute distances based on different metrics such as Euclidean, Manhattan, and Mahalanobis, the `dist` function offers flexibility in analyzing data. For instance, it can be used to group similar data points together in cluster analysis or to identify outliers in data. Additionally, the `dist` function can also be utilized in machine learning algorithms such as k-nearest neighbors (KNN) for classification tasks. Overall, the `dist` function is a powerful tool in R for measuring the similarity or dissimilarity between data points, making it a valuable tool for data analysis and decision making.

Use the dist Function in R (With Examples)

The `dist()` function in R can be used to calculate a distance matrix, which displays the distances between the rows of a matrix or data frame.

This function uses the following basic syntax:

```
dist(x, method="euclidean")
```

where:

x: The name of the matrix or data frame.
method: The distance measure to use. Default is "euclidean" but options include "maximum", "manhattan", "canberra", "binary" or "minkowski".

The following examples show how to use this function in practice with the following data frame:

#define four vectors

```
a <- c(2, 4, 4, 6)
```

```
b <- c(5, 5, 7, 8)
```

```
c <- c(9, 9, 9, 8)
```

```
d <- c(1, 2, 3, 3)
```

#row bind four vectors into matrix

```
mat <- rbind(a, b, c, d)
```

#view matrix

```
mat
```

```
a 2 4 4 6
```

```
b 5 5 7 8
```

```
c 9 9 9 8
```

```
d 1 2 3 3
```

Example 1: Use dist() to Calculate Euclidean Distance

The Euclidean distance between two vectors, A and B, is calculated as:

$$\text{Euclidean distance} = \sqrt{\sum(A_i - B_i)^2}$$

The following code shows how to compute a distance matrix that shows the Euclidean distance between each

row of a matrix in R:

#calculate Euclidean distance between each row in matrix

dist(mat)

a b c

b 4.795832

c 10.148892 6.000000

d 3.872983 8.124038 13.190906

Here's how to interpret the output:

The Euclidean distance between row a and row b is 4.795832. The Euclidean distance between row a and row c is 10.148892. The Euclidean distance between row a and row d is 3.872983. The Euclidean distance between row b and row c is 6.000000. The Euclidean distance between row b and row d is 8.124038. The Euclidean distance between row c and row d is 13.190906.

Example 2: Use dist() to Calculate Maximum Distance

The Maximum distance between two vectors, A and B, is calculated as the maximum difference between any

pairwise elements.

The following code shows how to compute a distance matrix that shows the Maximum distance between each row of a matrix in R:

```
#calculate Maximum distance between each row in matrix  
dist(mat, method="maximum")
```

```
a b c  
b 3  
c 7 4  
d 3 5 8
```

Example 3: Use dist() to Calculate Canberra Distance

The Canberra distance between two vectors, A and B, is calculated as:

Canberra distance = $\sum |A_i - B_i| / |A_i| + |B_i|$

The following code shows how to compute a distance matrix that shows the Canberra distance between each row of a matrix in R:

#calculate Canberra distance between each row in matrix

```
dist(mat, method="canberra")
```

```
a b c
```

```
b 0.9552670
```

```
c 1.5484515 0.6964286
```

```
d 1.1428571 1.9497835 2.3909091
```

Example 4: Use dist() to Calculate Binary Distance

The Binary distance between two vectors, A and B, is calculated as the proportion of elements that the two vectors share.

The following code shows how to compute a distance matrix that shows the Binary distance between each row of a matrix in R:

```
#calculate Binary distance between each row in matrix  
dist(mat, method="binary")
```

```
a b c
```

```
b 0
```

```
c 0 0
```

d 0 0 0

Example 5: Use dist() to Calculate Minkowski Distance

The Minkowski distance between two vectors, A and B, is calculated as:

$$\text{Minkowski distance} = (\sum |a_i - b_i|^p)^{1/p}$$

where i is the i th element in each vector and p is an integer.

The following code shows how to compute a distance matrix that shows the Minkowski distance (using $p=3$) between each row of a matrix in R:

```
#calculate Minkowski distance between each row in matrix  
dist(mat, method="minkowski", p=3)
```

a b c

b 3.979057

c 8.439010 5.142563

d 3.332222 6.542133 10.614765