

What is the process for normalizing data between 0 and 1?

Authored by
stats writer

April 29, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the process for normalizing data between 0 and 1?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=140911>

The process of normalizing data between 0 and 1 is a statistical technique used to standardize data values to a common scale. This is achieved by subtracting the minimum value from each data point and then dividing it by the range (maximum value minus minimum value). This results in all data points being scaled between 0 and 1, making it easier to compare and analyze different sets of data. Normalization helps to eliminate the influence of different units of measurement and allows for a more accurate comparison of data. This process is commonly used in data analysis, machine learning, and other statistical applications.

Normalize Data Between 0 and 1

To normalize the values in a dataset to be between 0 and 1, you can use the following formula:

$$z_i = (x_i - \min(x)) / (\max(x) - \min(x))$$

where:

z_i : The i th normalized value in the dataset
 x_i : The i th value in the dataset
 $\min(x)$: The minimum value in the dataset
 $\max(x)$: The maximum value in the dataset

For example, suppose we have the following dataset:

Data Values
13
16
19
22
23
38
47
56
58
63
65
70
71

The minimum value in the dataset is 13 and the maximum value is 71.

To normalize the first value of 13, we would apply the formula shared earlier:

$$z_i = (x_i - \min(x)) / (\max(x) - \min(x)) = (13 - 13) / (71 - 13) = 0$$

To normalize the second value of 16, we would use the same formula:

$$z_i = (x_i - \min(x)) / (\max(x) - \min(x)) = (16 - 13) / (71 - 13) = .0517$$

To normalize the third value of 19, we would use the same formula:

$$z_i = (x_i - \min(x)) / (\max(x) - \min(x)) = (19 - 13) / (71 - 13) = .1034$$

We can use this exact same formula to normalize each value in the original dataset to be between 0 and 1:

Data Values	Normalized
13	0
16	0.0517
19	0.1034
22	0.1552
23	0.1724
38	0.4310
47	0.5862
56	0.7414
58	0.7759
63	0.8621
65	0.8966
70	0.9828
71	1

Using this normalization method, the following statements will always be true:

The normalized value for the minimum value in the dataset will always be 0. The normalized value for the

maximum value in the dataset will always be 1. The normalized values for all other values in the dataset will be between 0 and 1.

When to Normalize Data

Often we normalize variables when performing some type of analysis in which we have multiple variables that are measured on different scales and we want each of the variables to have the same range.

This prevents one variable from being overly influential, especially if it's measured in different units (i.e. if one variable is measured in inches and another is measured in yards).

It's also worth noting that we used a method known as min-max normalization in this tutorial to normalize the data values.

The two most common normalization methods are as follows:

1. Min-Max Normalization

Objective: Converts each data value to a value between

0 and 100. Formula: $\text{New value} = (\text{value} - \text{min}) / (\text{max} - \text{min}) * 100$

2. Mean Normalization

Objective: Scales values such that the mean of all values is 0 and std. dev. is 1. Formula: $\text{New value} = (\text{value} - \text{mean}) / (\text{standard deviation})$

The following tutorials explain how to normalize data using different statistical softwares:

How to Normalize Data in Excel

How to Normalize Data in R

How to Normalize Columns in Python