

What is the difference between standardization and normalization?

Authored by
stats writer

May 1, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the difference between standardization and normalization?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=141763>

Standardization and normalization are two commonly used methods in data preprocessing, which involves transforming raw data into a more usable format for analysis. While both methods aim to make data more consistent and comparable, they have distinct differences.

Standardization is the process of transforming data to have a mean of 0 and a standard deviation of 1. This ensures that the data is centered around 0 and has a consistent spread. This method is useful when the data has a wide range of values and different units of measurement.

On the other hand, normalization is the process of scaling data to a range of 0 to 1. This is achieved by subtracting the minimum value from each data point and then dividing it by the range of values. This method is useful when the data has a skewed distribution or outliers.

In summary, standardization is useful for comparing data with different scales, while normalization is useful for comparing data with different distributions. Both methods are important in data preprocessing and the choice between them depends on the specific data and the goal of the analysis.

Standardization vs. Normalization: What's the Difference?

Standardization and normalization are two ways to rescale data.

Standardization rescales a dataset to have a mean of 0 and a standard deviation of 1. It uses the following formula to do so:

$$x_{\text{new}} = (x_i - \bar{x}) / s$$

where:

x_i : The i th value in the dataset
 $\bar{x}$: The sample means:

The sample standard deviation

Normalization rescales a dataset so that each value falls between 0 and 1. It uses the following formula to do so:

$$x_{\text{new}} = (x_i - x_{\text{min}}) / (x_{\text{max}} - x_{\text{min}})$$

where:

x_i : The i th value in the dataset
 x_{min} : The minimum value in the dataset
 x_{max} : The maximum value in the dataset

The following examples show how to standardize and normalize a dataset in practice.

Example: How to Standardize Data

Suppose we have the following dataset:

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

The mean value in the dataset is 43.15 and the standard deviation is 22.13.

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

$$x_{\text{new}} = (x_i - x) / s = (13 - 43.15) / 22.13 = -1.36$$

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

$$x_{\text{new}} = (x_i - x) / s = (16 - 43.15) / 22.13 = -1.23$$

To normalize the third value of 19, we would use the

same formula:

$$x_{\text{new}} = (x_i - \bar{x}) / s = (19 - 43.15) / 22.13 = -1.09$$

We can use this exact same formula to standardize each value in the original dataset:

Data Values	Standardized
13	-1.36
16	-1.23
19	-1.09
22	-0.96
23	-0.91
38	-0.23
47	0.17
56	0.58
58	0.67
63	0.90
65	0.99
70	1.21
71	1.26

Example: How to Normalize Data

Once again, 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:

$$x_{\text{new}} = (x_i - x_{\text{min}}) / (x_{\text{max}} - x_{\text{min}}) = (13 - 13) / (71 - 13) = 0$$

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

$$x_{\text{new}} = (x_i - x_{\text{min}}) / (x_{\text{max}} - x_{\text{min}}) = (16 - 13) / (71 - 13) = .0517$$

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

$$x_{\text{new}} = (x_i - x_{\text{min}}) / (x_{\text{max}} - x_{\text{min}}) = (19 - 13) / (71 - 13) = .1034$$

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

Standardization vs. Normalization: When to Use Each

Typically we normalize data 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).

On the other hand, we typically standardize data when we'd like to know how many standard deviations each value in a dataset lies from the mean.

For example, we might have a list of exam scores for 500 students at a particular school and we'd like to know how many standard deviations each exam score lies from the mean score.

In this case, we could standardize the raw data to find out this information. Then, a standardized score of 1.26 would tell us that the exam score of that particular student lies 1.26 standard deviations above the mean exam score.

Whether you decide to normalize or standardize your data, keep the following in mind:

A normalized dataset will always have values that range between 0 and 1. A standardized dataset will have a

mean of 0 and standard deviation of 1, but there is no specific upper or lower bound for the maximum and minimum values.

Depending on your particular scenario, it may make more sense to normalize or standardize the data.

The following tutorials explain how to standardized and normalize data in different statistical software:

ARABPSYCHOLOGY.COM