

What is the definition of Z-score normalization and what are some examples?

Authored by
stats writer

May 5, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the definition of Z-score normalization and what are some examples?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=142973>

Z-score normalization, also known as standardization, is a statistical technique used to transform data into a standard normal distribution. This method is commonly used in data analysis and machine learning to compare and interpret different datasets. It involves calculating the mean and standard deviation of a dataset and then transforming the data points to have a mean of 0 and a standard deviation of 1. This allows for easier comparison and interpretation of data. For example, if a dataset contains the heights of individuals in inches and another dataset contains the heights in centimeters, the two datasets can be standardized using Z-score normalization to compare them accurately. Other examples include transforming test scores, salaries, and stock prices.

Z-Score Normalization: Definition & Examples

Z-score normalization refers to the process of normalizing every value in a dataset such that the mean of all of the values is 0 and the standard deviation is 1.

We use the following formula to perform a z-score normalization on every value in a dataset:

$$\text{New value} = (x - \mu) / \sigma$$

where:

x: Original value **μ : Mean of data** **σ : Standard deviation of data**

The following example shows how to perform z-score normalization on a dataset in practice.

Example: Performing Z-Score Normalization

Suppose we have the following dataset:

Data
3
5
5
8
9
12
12
13
15
16
17
19
22
24
25
134

Using a calculator, we can find that the mean of the dataset is 21.2 and the standard deviation is 29.8.

To perform a z-score normalization on the first value in the dataset, we can use the following formula:

$$\text{New value} = (x - \mu) / \sigma$$
$$\text{New value} = (3 - 21.2) / 29.8$$
$$\text{New value} = -0.61$$

We can use this formula to perform a z-score normalization on every value in the dataset:

Data	Z-Score Normalized Value
3	-0.61
5	-0.54
5	-0.54
8	-0.44
9	-0.41
12	-0.31
12	-0.31
13	-0.28
15	-0.21
16	-0.17
17	-0.14
19	-0.07
22	0.03
24	0.09
25	0.13
134	3.79

The mean of the normalized values is 0 and the standard deviation of the normalized values is 1.

The normalized values represent the number of standard deviations that the original value is from the mean.

For example:

The first value in the dataset is 0.61 standard deviations below the mean. The second value in the dataset is 0.54 standard deviations below the mean.... The last value in the dataset is 3.79 standard deviations above the mean.

The benefit of performing this type of normalization is that the clear outlier in the dataset (134) has been transformed in such a way that it's no longer a massive outlier.

If we then use this dataset to fit some type of , the outlier will no longer have as big of an influence that it might have on the model fit.

The following tutorials provide additional information on different normalization techniques: