How can data be normalized in SAS?

Authored by

stats writer

July 1, 2024

Data normalization in SAS is a process of organizing and structuring data in a standardized and consistent manner, making it easier to analyze and compare. This involves identifying and removing any inconsistencies, redundancies, or errors in the data and transforming it into a uniform format. This can be achieved through SAS procedures such as PROC STANDARD, which allows for the scaling and shifting of data to a common mean and standard deviation, and PROC SQL, which enables the merging and combining of data from different sources. Normalizing data in SAS helps to improve data quality, enhance data analysis, and facilitate efficient decision making.

Normalize Data in SAS

To "normalize" a set of data values means to scale the values such that the mean of all of the values is 0 and the standard deviation is 1.

This tutorial explains how to normalize data in SAS.

Example: How to Normalize Data in SAS

Suppose we have the following dataset:

Data
12
14
15
15
16
17
18
20
24
25
26
29
32
34
37

Perform the following steps to normalize this set of data values in SAS.

Step 1: Create the Dataset

First, let's use the following code to create the dataset in SAS:

```
/*create dataset*/
data original_data;
input values;
datalines;
12
```

```
14
15
15
16
17
18
20
24
25
26
29
32
34
37
;
run;

/*view mean and standard deviation of dataset*/
proc meansdata=original_data Mean StdDev ndec=3;
var values;
run;
```

The MEANS Procedure

Analysis Variable : values	
Mean	Std Dev
22.267	7.968

From the output we can see that the mean of the dataset is 22.267 and the standard deviation is 7.968.

Step 2: Normalize the Dataset

Next, we'll use proc stdize to normalize the dataset:

```
/*normalize the dataset*/
proc stdizedata=original_data out=normalized_data;
var values;
run;

/*print normalized dataset*/
proc printdata=normalized_data;

/*view mean and standard deviation of normalized dataset*/
proc meansdata=normalized_data MeanStdDev ndec=2;
var values;
run;
```

Obs	values
1	-1.28842
2	-1.03743
3	-0.91194
4	-0.91194
5	-0.78644
6	-0.66094
7	-0.53545
8	-0.28446
9	0.21753
10	0.34302
11	0.46852
12	0.84501
13	1.22149
14	1.47248
15	1.84897

The MEANS Procedure

Analysis Variable : values	
Mean	Std Dev
0.00	1.00

From the output we can see that the mean of the normalized dataset is 0 and the standard deviation is 1.

Step 3: Interpret the Normalized Data

Normalized value = (x - x) / s

where:

x = data valuex = mean of datasets = standard deviation of dataset

Each normalized value tells us how many standard deviations the original data value was from the mean.

For example, consider the data point "12" in our original dataset. The original sample mean was 22.267 and the original sample standard deviation was 7.968.

The normalized value for "12" turned out to be -1.288, which was calculated as:

Normalized value = (x - x) / s = (12 - 22.267) / 7.968 = -1.288

This tells us that the value "12" is 1.288 standard deviations below the mean in the original dataset.

Each of the normalized values in the dataset can help us understand how close or far a particular data value is from the mean.

A small normalized value indicates that a value is close to the mean while a large normalized value indicates that a value is far from the mean.

Additional Resources

The following tutorials explain how to perform other common tasks in SAS: