How can I calculate Z-scores in SAS?

Authored by

stats writer

June 26, 2024

Z-scores, also known as standard scores, are a statistical measure that indicates how many standard deviations a data point is above or below the mean of a data set. In order to calculate Z-scores in SAS, you can use the PROC STANDARD procedure. This procedure allows you to standardize your data by subtracting the mean and dividing by the standard deviation. The resulting Z-scores can then be used to compare data points from different distributions and identify outliers. By using the PROC STANDARD procedure, you can easily and accurately calculate Z-scores in SAS for your data analysis needs.

Calculate Z-Scores in SAS

In statistics, a z-score tells us how many standard deviations away a value is from .

We use the following formula to calculate a z-score:

$z = (X - \mu) / \sigma$

where:

X is a single raw data valueμ is the mean of the datasetσ is the standard deviation of the dataset

The following example shows how to calculate z-scores for raw data values in SAS.

Example: Calculate Z-Scores in SAS

Suppose we create the following dataset in SAS:

/*create dataset*/

```
data original_data;
input values;
datalines;
7
12
14
12
16
18
6
7
14
17
19
22
24
13
17
12
;
run;

/*view dataset*/
proc printdata=original_data;
```

Obs	values
1	7
2	12
3	14
4	12
5	16
6	18
7	6
8	7
9	14
10	17
11	19
12	22
13	24
14	13
15	17
16	12

Now suppose we would like to calculate the z-score for each value in the dataset.

We can use proc sql to do so:

```
/*create new variable that shows z-scores for each raw data value*/
proc sql;
select values, (values - mean(values)) / std(values) as z_scores
from original_data;
```

```
quit;
```

values	z_scores
7	-1.42861
12	-0.46006
14	-0.07264
12	-0.46006
16	0.314778
18	0.702198
6	-1.62232
7	-1.42861
14	-0.07264
17	0.508488
19	0.895907
22	1.477037
24	1.864456
13	-0.26635
17	0.508488
12	-0.46006

The values column shows the original data values and the z_scores column shows the z-score for each value.

How to Interpret Z-Scores in SAS

A z-score tells us how many standard deviations away a value is from the mean.

A positive z-score indicates that a particular value is

greater than the mean, a negative z-score indicates that a particular value is less than the mean, and a z-score of zero indicates that a particular value is equal to the mean.

If we calculated the mean and standard deviation of our dataset, we would find that the mean is 14.375 and the standard deviation is 5.162.

So, the first value in our dataset was 7, which had a z-score of (7-14.375) / 5.162 = -1.428. This means that the value "7" is 1.428 standard deviations *below* the mean.

The next value in our data, 12, had a z-score of (12-14.375) / 5.162 = -0.46. This means that the value "12" is 0.46 standard deviations *below* the mean.

The further away a value is from the mean, the higher the absolute value of the z-score will be for that value.

For example, the value 7 is further away from the mean (14.375) compared to 12, which explains why 7 had a z-score with a larger absolute value.

The following articles explain how to perform other

common tasks in SAS: