

How to use PROC STDIZE in SAS with an example?

Authored by
stats writer

June 23, 2024

RECOMMENDED CITATION

stats writer (2024). *How to use PROC STDIZE in SAS with an example?*.
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=147819>

PROC STDIZE is a SAS procedure that allows users to standardize numerical variables, also known as z-score transformation, in order to make them comparable and easier to analyze. This procedure is particularly useful when dealing with data that has a wide range of values and helps to remove any potential biases caused by extreme values.

To use PROC STDIZE in SAS, the first step is to specify the variables that need to be standardized in the VAR statement. Users can choose to standardize all variables or specific ones by listing their names. Next, the METHOD statement should be used to select the type of standardization desired, such as mean-centered or range-standardized. Finally, the OUTPUT statement can be used to create a new dataset with the standardized variables.

For example, if we have a dataset with three numerical variables, "height," "weight," and "age," and we want to standardize all three, we can use the following code:

```
PROC STDIZE DATA = original_data OUT = standardized_data;  
VAR height weight age;  
METHOD = MEAN;  
RUN;
```

This will create a new dataset called "standardized_data" with the three variables standardized using the mean-centered method. This allows us to compare the values of each variable on the same scale, making it easier to identify patterns and relationships in the data.

In conclusion, PROC STDIZE in SAS is a powerful tool for standardizing numerical variables and improving the accuracy and reliability of data analysis. It is essential for any researcher or analyst working with large and diverse datasets.

Use PROC STDIZE in SAS (With Example)

To standardize a variable means to scale each of the values for the variable such that the mean value is 0 and the standard deviation is 1.

You can use the following formula to standardize a variable

$$(x_i - \bar{x}) / s$$

where:

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

The easiest way to standardize a variable in SAS is to use the PROC STDIZE statement.

The following example shows how to use this statement in practice.

Example: How to Use PROC STDIZE in SAS

Suppose we have the following dataset in SAS that contains information about various basketball players:

```
/*create first dataset*/  
data my_data;  
input player $ points assists rebounds;  
datalines;  
A 18 3 15  
B 20 3 14  
C 19 4 14  
D 14 5 10
```

```
E 14 4 8  
F 15 7 14  
G 20 8 13  
H 28 7 9  
I 30 6 5  
J 0 31 9 4  
;  
run;
```

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

Obs	player	points	assists	rebounds
1	A	18	3	15
2	B	20	3	14
3	C	19	4	14
4	D	14	5	10
5	E	14	4	8
6	F	15	7	14
7	G	20	8	13
8	H	28	7	9
9	I	30	6	5
10	J	0	31	9

We can use the PROC STDIZE statement to create a new dataset that standardizes each of the numeric

variables in the dataset:

```
/*standardize all numeric variables in dataset*/
```

```
proc stdizedata=my_data out=std_data;
```

```
run;
```

```
/*view new dataset*/
```

```
proc printdata=std_data;
```

Obs	player	points	assists	rebounds
1	A	0.02414	-0.57572	1.16477
2	B	0.26558	-0.57572	0.86611
3	C	0.14486	-0.45578	0.86611
4	D	-0.45872	-0.33584	-0.32852
5	E	-0.45872	-0.45578	-0.92584
6	F	-0.33801	-0.09595	0.86611
7	G	0.26558	0.02399	0.56745
8	H	1.23131	-0.09595	-0.62718
9	I	1.47274	-0.21590	-1.82182
10	J	-2.14876	2.78266	-0.62718

Each of the numeric variables (points, assists, rebounds) have been standardized to have a mean of 0 and standard deviation of 1.

Note that we can also use the VAR statement to specify which variables to standardize.

For example, we can use the following PROC STDIZE statement with the VAR statement to only standardize the points variable:

```
/*standardize points variable in dataset*/  
proc stdizedata=my_data out=std_data;  
var points;  
run;
```

```
/*view new dataset*/  
proc printdata=std_data;
```

Obs	player	points	assists	rebounds
1	A	0.02414	3	15
2	B	0.26558	3	14
3	C	0.14486	4	14
4	D	-0.45872	5	10
5	E	-0.45872	4	8
6	F	-0.33801	7	14
7	G	0.26558	8	13
8	H	1.23131	7	9
9	I	1.47274	6	5
10	J	-2.14876	31	9

The values in the points column have been standardized while all other columns have remained untouched.

We can use the PROC MEANS statement to verify that the points variable now has a mean value of 0 and a standard deviation of 1:

```
/*view mean and standard deviation of each variable*/proc meansdata=std_data;
```

The MEANS Procedure					
Variable	N	Mean	Std Dev	Minimum	Maximum
points	10	-4.44089E-17	1.0000000	-2.1487585	1.4727446
assists	10	7.8000000	8.3373324	3.0000000	31.0000000
rebounds	10	11.1000000	3.3482997	5.0000000	15.0000000

We can see that the points variable indeed has a mean value of 0 and a standard deviation of 1.

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