How can a log transformation be performed in SAS?

Authored by

stats writer

June 23, 2024

A log transformation in SAS is a statistical technique used to transform data from its original scale to a logarithmic scale. This transformation is often applied to data that is skewed or has a wide range of values. To perform a log transformation in SAS, the LOG function can be used. This function takes the natural logarithm of a variable and creates a new variable with the transformed values. The LOG function can be applied in various SAS procedures, such as PROC TRANSPOSE or PROC UNIVARIATE, to transform individual variables or entire datasets. Additionally, the LOG10 function can be used to perform a base 10 logarithmic transformation. This technique can help to normalize data and improve the accuracy of statistical analyses.

Perform a Log Transformation in SAS

Many statistical tests make the assumption that the values for a particular variable are .

However, often values are *not* normally distributed. One way to address this issue is to transform the variable by taking the log of each value.

By performing this transformation, a variable typically becomes closer to normally distributed.

The following example shows how to perform a log transformation on a variable in SAS.

Example: Log Transformation in SAS

Suppose we have the following dataset in SAS:

/*create dataset*/

```
data my_data;
input x;
datalines;
1
1
1
2
2
2
2
2
2
3
3
3
6
7
8
;
run;

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

Obs	x
1	1
2	1
3	1
4	2
5	2
6	2
7	2
8	2
9	2
10	3
11	3
12	3
13	6
14	7
15	8

We can use to perform normality tests on the variable x to determine if it is normally distributed and also create a histogram to visualize the distribution of values:

```
/*create histogram and perform normality tests*/
proc univariatedata=my_data normal;
histogram x;
run;
```

The UNIVARIATE Procedure
Variable: x

Moments			
N	15	Sum Weights	15
Mean	3	Sum Observations	45
Std Deviation	2.20389266	Variance	4.85714286
Skewness	1.43206094	Kurtosis	0.96326857
Uncorrected SS	203	Corrected SS	68
Coeff Variation	73.4630887	Std Error Mean	0.56904264

Basic Statistical Measures			
Location		Variability	
Mean	3.000000	Std Deviation	2.20389
Median	2.000000	Variance	4.85714
Mode	2.000000	Range	7.00000
		Interquartile Range	1.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	5.272013	Pr > |t|	0.0001
Sign	M	7.5	Pr >= |M|	<.0001
Signed Rank	S	60	Pr >= |S|	<.0001

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.772253	Pr < W	0.0017
Kolmogorov-Smirnov	D	0.3	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.286181	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	1.512189	Pr > A-Sq	<0.0050

From the last table titled Tests for Normality we can see that the for the Shapiro-Wilk test is less than .05, which provides strong evidence that the variable x is not

normally distributed.

The histogram also shows that the distribution of values does not appear to be normally distributed:

We can attempt a log transformation on the original dataset to see if we can produce a dataset that is more normally distributed.

We can use the following code to create a new dataset in SAS in which we take the log of each of the original x values:

```
/*use log transformation to create new dataset*/
data log_data;
set my_data;
x = log(x);
run;

/*view log transformed data*/
proc printdata=log_data;
```

Obs	x
1	0.00000
2	0.00000
3	0.00000
4	0.69315
5	0.69315
6	0.69315
7	0.69315
8	0.69315
9	0.69315
10	1.09861
11	1.09861
12	1.09861
13	1.79176
14	1.94591
15	2.07944

```
/*create histogram and perform normality tests*/
proc univariatedata=log_data normal;
histogram x;
run;
```

The UNIVARIATE Procedure
Variable: x

Moments			
N	15	Sum Weights	15
Mean	0.88478874	Sum Observations	13.2718311
Std Deviation	0.65910349	Variance	0.4344174
Skewness	0.44759883	Kurtosis	-0.3719325
Uncorrected SS	17.8246104	Corrected SS	6.08184366
Coeff Variation	74.4927523	Std Error Mean	0.17017979

Basic Statistical Measures			
Location		Variability	
Mean	0.884789	Std Deviation	0.65910
Median	0.693147	Variance	0.43442
Mode	0.693147	Range	2.07944
		Interquartile Range	0.40547

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	5.199141	Pr > |t|	0.0001
Sign	M	6	Pr >= |M|	0.0005
Signed Rank	S	39	Pr >= |S|	0.0005

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.890891	Pr < W	0.0692
Kolmogorov-Smirnov	D	0.214383	Pr > D	0.0630
Cramer-von Mises	W-Sq	0.127487	Pr > W-Sq	0.0435
Anderson-Darling	A-Sq	0.718565	Pr > A-Sq	0.0480

From the last table titled Tests for Normality we can see that the for the Shapiro-Wilk test is now greater than .05.

The histogram also shows that the distribution of values is slightly more normally distributed than it was before the transformation:

Based on the results of the Shapiro-Wilk test and the histogram shown above, we would conclude that the log transformation created a variable that is much more normally distributed than the original variable.

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