

How can I perform a Box-Cox transformation in Excel?

Authored by
stats writer

April 28, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I perform a Box-Cox transformation in Excel?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=140690>

Box-Cox transformation is a statistical technique used to transform non-normally distributed data into a normal distribution. This transformation is useful for improving the accuracy and interpretability of statistical models. To perform a Box-Cox transformation in Excel, you can use the built-in function called "POWER" and the "LOG" function. This involves taking the logarithm of the data, raising it to a power, and then applying the inverse of the logarithm. By using this method, you can transform your data to meet the assumptions of normality and improve the results of your analysis.

Box-Cox Transformation in Excel (Step-by-Step)

A box-cox transformation is a commonly used method for transforming a non-normally distributed dataset into a more normally distributed one.

The basic idea is to find some value for λ such that the transformed data is as close to normally distributed as possible, using the following formula:

$$y(\lambda) = (y^\lambda - 1) / \lambda \text{ if } y \neq 0 \quad y(\lambda) = \log(y) \text{ if } y = 0$$

The following step-by-step example shows how to perform a box-cox transformation on a dataset in Excel.

Step 1: Enter the Data

First, let's enter the values for a dataset:

	A	B	C	D	E	F
1	Raw Data					
2	4					
3	5					
4	2					
5	3					
6	3					
7	2					
8	2					
9	3					
10	2					
11	2					
12	3					
13	4					
14	3					
15	8					
16	6					
17						
18						
19						
20						
21						
22						
23						
24						

Step 2: Sort the Data

Next, create an index column and a column of sorted data:

	A	B	C	D	E	F	G
1	Raw Data	Index	Sorted				
2	4	1	2	=SMALL(\$A\$2:\$A\$16, B2)			
3	5	2	2				
4	2	3	2				
5	3	4	2				
6	3	5	2				
7	2	6	3				
8	2	7	3				
9	3	8	3				
10	2	9	3				
11	2	10	3				
12	3	11	4				
13	4	12	4				
14	3	13	5				
15	8	14	6				
16	6	15	8				
17							
18							
19							
20							
21							
22							
23							
24							

Step 3: Choose an Arbitrary Value for Lambda

Next, we'll choose an arbitrary value of 1 for lambda and apply a temporary box-cox transformation to the data:

	A	B	C	D	E	F	G	H
1	Raw Data	Index	Sorted	(Sorted ^λ -1)/λ			λ	1
2	4	1	2	1	= (C2^H\$1-1)/H\$1			
3	5	2	2	1				
4	2	3	2	1				
5	3	4	2	1				
6	3	5	2	1				
7	2	6	3	2				
8	2	7	3	2				
9	3	8	3	2				
10	2	9	3	2				
11	2	10	3	2				
12	3	11	4	3				
13	4	12	4	3				
14	3	13	5	4				
15	8	14	6	5				
16	6	15	8	7				
17								
18								
19								
20								

Step 4: Calculate the Z-Scores

Next, we'll calculate the z-score for each value in the index:

	A	B	C	D	E	F	G	H
1	Raw Data	Index	Sorted	$(\text{Sorted}^\lambda - 1)/\lambda$	z	λ		1
2	4	1	2	1	-1.834	=NORM.S.INV((B2-0.5)/\$B\$16)		
3	5	2	2	1	-1.282			
4	2	3	2	1	-0.967			
5	3	4	2	1	-0.728			
6	3	5	2	1	-0.524			
7	2	6	3	2	-0.341			
8	2	7	3	2	-0.168			
9	3	8	3	2	0.000			
10	2	9	3	2	0.168			
11	2	10	3	2	0.341			
12	3	11	4	3	0.524			
13	4	12	4	3	0.728			
14	3	13	5	4	0.967			
15	8	14	6	5	1.282			
16	6	15	8	7	1.834			
17								
18								
19								
20								

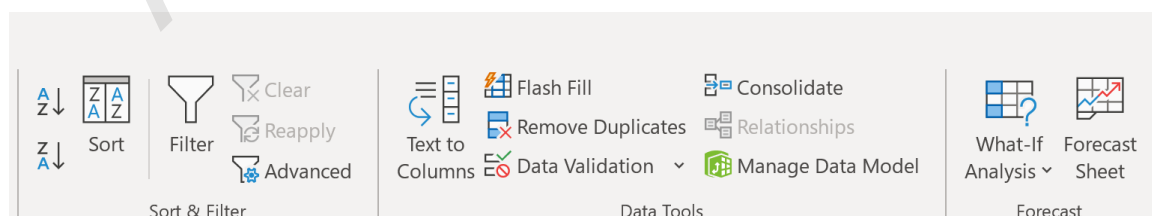
We'll then calculate the correlation between the box-cox transformed values and the z-scores:

	A	B	C	D	E	F	G	H	I	J	K
1	Raw Data	Index	Sorted	$(\text{Sorted}^\lambda - 1)/\lambda$	z		λ	1			
2	4	1	2	1	-1.834		r	0.89803	=CORREL(D2:D16, E2:E16)		
3	5	2	2	1	-1.282						
4	2	3	2	1	-0.967						
5	3	4	2	1	-0.728						
6	3	5	2	1	-0.524						
7	2	6	3	2	-0.341						
8	2	7	3	2	-0.168						
9	3	8	3	2	0.000						
10	2	9	3	2	0.168						
11	2	10	3	2	0.341						
12	3	11	4	3	0.524						
13	4	12	4	3	0.728						
14	3	13	5	4	0.967						
15	8	14	6	5	1.282						
16	6	15	8	7	1.834						
17											
18											
19											
20											
21											
22											
23											

Step 5: Find the Optimal Lambda Value

Next, we'll use Goal Seek to find the optimal lambda value to use in the box-cox transformation.

To do so, click the Data tab along the top ribbon. Then click What-If-Analysis within the Forecast group.



In the dropdown menu, click Goal Seek and fill in the

following values:

	A	B	C	D	E	F	G	H	I
1	Raw Data	Index	Sorted	$(\text{Sorted}^{\lambda}-1)/\lambda$	z		λ		
2	4	1	2	1	-1.834		r	0.89803	
3	5	2	2	1	-1.282				
4	2	3	2	1	-0.967				
5	3	4	2	1	-0.728				
6	3	5	2	1	-0.524				
7	2	6	3	2	-0.341				
8	2	7	3	2	-0.168				
9	3	8	3	2	0.000				
10	2	9	3	2	0.168				
11	2	10	3	2	0.341				
12	3	11	4	3	0.524				
13	4	12	4	3	0.728				
14	3	13	5	4	0.967				
15	8	14	6	5	1.282				
16	6	15	8	7	1.834				
17									
18									
19									
20									
21									

Goal Seek ? X

Set cell: ↑

To value:

By changing cell: ↑

OK Cancel

Once you click OK, Goal Seek will automatically find the optimal lambda value to be -0.5225.

E	F	G	H	I	J
z		λ	-0.5225		
834		r	0.948791		
282					
967					
728					
524					
341					
168					
000					
168					
341					
524					
728					
967					
282					
834					

Goal Seek Status ? X

Goal Seeking with Cell H2
may not have found a solution.

Target value: 1

Current value: 0.948790681

Step

Pause

OK

Cancel

Step 6: Perform the Box-Cox Transformation

Lastly, we'll apply the box-cox transformation to the original data, using a lambda value of -0.5225:

	A	B	C	D	E	F	G	H	I
1	Raw Data	Index	Sorted	$(\text{Sorted}^{\lambda}-1)/\lambda$	z	Transformed Data			
2	4	1	2	1	-1.834	0.986	$=(A2^{(-0.5225)}-1)/-0.5225$		
3	5	2	2	1	-1.282	1.088			
4	2	3	2	1	-0.967	0.582			
5	3	4	2	1	-0.728	0.836			
6	3	5	2	1	-0.524	0.836			
7	2	6	3	2	-0.341	0.582			
8	2	7	3	2	-0.168	0.582			
9	3	8	3	2	0.000	0.836			
10	2	9	3	2	0.168	0.582			
11	2	10	3	2	0.341	0.582			
12	3	11	4	3	0.524	0.836			
13	4	12	4	3	0.728	0.986			
14	3	13	5	4	0.967	0.836			
15	8	14	6	5	1.282	1.268			
16	6	15	8	7	1.834	1.163			
17									
18									
19									
20									
21									
22									
23									
24									

Bonus: We can confirm that the transformed data is normally distributed by performing a .