How can I perform bootstrapping in Excel, and what is an example of how it can be used?

Authored by

stats writer

June 28, 2024

Bootstrapping is a statistical method used to estimate the accuracy of a sample statistic by creating a large number of samples from a single dataset. In Excel, bootstrapping can be performed by using the "Data Analysis" tool, which allows users to generate multiple random samples from a given dataset.

One example of how bootstrapping can be used in Excel is to estimate the mean of a population when the sample size is small. By generating multiple samples from the original dataset and calculating the means of each sample, a distribution of means can be created. This distribution can then be used to estimate the confidence interval for the population mean and provide a more accurate representation of the true population mean. This method is particularly useful when the sample size is small and a traditional statistical test may not be applicable.

Perform Bootstrapping in Excel (With Example)

Bootstrapping is a method that can be used to construct a confidence interval for a when the sample size is small and the underlying distribution is unknown.

The basic process for bootstrapping is as follows:

Take *k* repeated samples with replacement from a given dataset.For each sample, calculate the statistic you're interested in.This results in *k* different estimates for a given statistic, which you can then use to calculate a confidence interval for the statistic.

The following step-by-step example shows how to perform bootstrapping in Excel.

Step 1: Enter the Original Data

First, we'll enter the values for some dataset:

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

Step 2: Generate Bootstrap Samples

Next, we'll use the following formula to generate bootstrap samples:

=INDEX($A$2:$A$16, RANDBETWEEN(1, ROWS($A$2:$A$16)),1)

We can type this formula into cell D2 to randomly select one value from the original dataset.

We can then drag this formula to the right for 10 cells to generate our first bootstrapped sample.

We can then drag this formula down 300 rows to create 300 bootstrapped samples:

D2 =INDEX($A$2:$A$16, RANDBETWEEN(1, ROWS($A$2:$A$16)),1)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	**Original Data**		**Boostrap Samples**											
2	7		Sample #1	19	12	10	12	17	15	15	9	12	17	
3	9		Sample #2	20	12	15	16	19	10	16	16	7	19	
4	10		Sample #3	10	21	14	10	21	10	16	9	10	20	
5	10		Sample #4	7	9	16	20	23	21	7	14	19	15	
6	12		Sample #5	15	10	10	20	10	21	14	20	20	23	
7	14		Sample #6	20	14	20	19	23	14	15	14	15	21	
8	15		Sample #7	12	14	12	19	15	14	21	16	9	17	
9	16		Sample #8	16	23	7	23	23	9	10	19	21	20	
10	16		Sample #9	10	23	20	10	21	7	9	16	19	19	
11	17		Sample #10	16	10	16	7	12	21	14	23	15	17	
12	19		Sample #11	16	7	10	19	16	16	16	10	15	15	
13	20		Sample #12	9	10	16	10	10	15	14	7	10	19	
14	21		Sample #13	17	10	12	21	12	23	12	20	16	16	
15	21		Sample #14	9	9	7	9	21	19	16	19	21	21	
16	23		Sample #15	19	10	15	9	21	14	7	16	16	17	
17			Sample #16	17	10	21	10	10	7	16	16	10	10	
18			Sample #17	15	21	9	17	10	12	16	14	9	16	
19			Sample #18	20	16	19	16	21	10	10	20	21	10	
20			Sample #19	23	7	16	19	21	17	7	10	16	10	
21			Sample #20	20	20	21	14	21	19	16	17	20	14	
22			Sample #21	23	23	19	12	9	23	16	17	19	16	
23			Sample #22	10	10	12	12	21	19	21	23	7	21	
24			Sample #23	21	21	16	10	15	16	16	16	15	10	
25			Sample #24	17	10	19	10	15	21	14	10	9	21	
26			Sample #25	17	17	12	7	16	21	10	12	19	7	
27			Sample #26	14	16	16	16	10	7	14	15	7	10	
28			Sample #27	7	16	19	21	10	17	15	21	7	14	
29			Sample #28	17	16	14	21	14	21	21	20	21	17	
30			Sample #29	15	21	15	21	7	20	21	14	10	23	
31			Sample #30	23	10	23	12	14	17	10	17	10	12	
32			Sample #31	15	21	23	10	17	9	10	7	10	19	
33			Sample #32	23	20	14	9	12	23	20	12	21	10	

Sheet1

Note: Bootstrapping uses , which means that one value from the original dataset may appear multiple times in any given sample.

Step 3: Calculate Statistic of Interest for Each Sample

Next, we can calculate the statistic of interest for each sample.

For this particular example, we'll calculate the median value for each sample:

N2 =MEDIAN(D2:M2)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Original Data		Boostrap Samples											Median
2	7		Sample #1	16	20	12	14	14	7	14	14	21	10	14
3	9		Sample #2	16	9	12	21	16	10	16	15	23	19	16
4	10		Sample #3	9	21	21	16	15	7	23	12	10	10	13.5
5	10		Sample #4	9	20	20	10	9	16	16	20	7	21	16
6	12		Sample #5	23	21	16	16	12	16	10	23	21	12	16
7	14		Sample #6	15	16	14	12	16	21	23	19	17	15	16
8	15		Sample #7	10	16	21	14	16	15	16	15	21	10	15.5
9	16		Sample #8	21	23	17	12	21	16	14	15	17	15	16.5
10	16		Sample #9	9	12	21	9	15	14	16	17	10	21	14.5
11	17		Sample #10	10	19	10	20	10	14	16	12	16	15	14.5
12	19		Sample #11	20	15	21	16	9	12	17	19	20	19	18
13	20		Sample #12	21	16	21	16	7	14	9	20	15	7	15.5
14	21		Sample #13	14	16	15	7	20	14	21	16	20	23	16
15	21		Sample #14	16	21	10	9	12	21	20	21	14	17	16.5
16	23		Sample #15	7	21	19	17	19	21	14	16	19	19	19
17			Sample #16	17	19	23	23	20	14	16	20	17	21	19.5
18			Sample #17	7	21	14	17	16	12	19	15	10	17	15.5
19			Sample #18	15	12	23	23	17	12	12	15	14	20	15
20			Sample #19	10	19	23	9	23	16	15	19	21	7	17.5
21			Sample #20	12	15	14	15	10	10	10	21	15	21	14.5
22			Sample #21	21	9	21	12	10	15	23	14	10	9	13
23			Sample #22	9	23	20	14	10	12	21	21	17	7	15.5
24			Sample #23	14	20	21	7	21	21	17	19	21	10	19.5
25			Sample #24	17	14	14	16	10	9	12	23	17	12	14
26			Sample #25	15	9	19	9	21	15	10	16	10	15	15

We can see:

The first bootstrapped sample has a median value of 14.The second bootstrapped sample has a median

value of 16.The third bootstrapped sample has a median value of 13.5.

And so on.

Step 4: Calculate Bootstrapped Confidence Interval

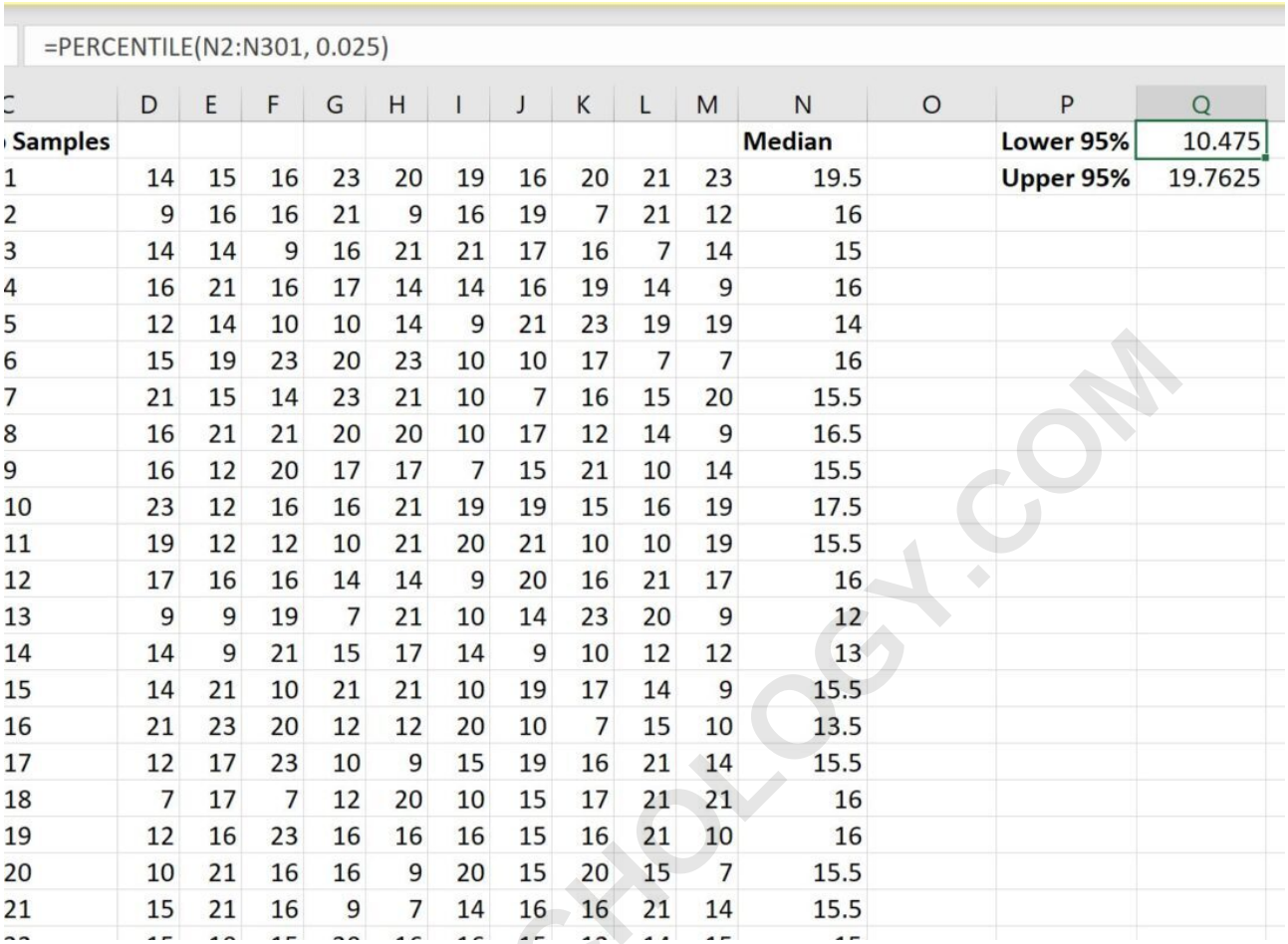
Lastly, we can calculate a 95% bootstrapped confidence interval for the median by finding the value located at percentile 2.5% and percentile 97.5% in column N.

We can use the following formulas to do so:

=PERCENTILE(N2:N301, 0.025)

=PERCENTILE(N2:N301, 0.975)

The following screenshot shows how to use these formulas in practice:



=PERCENTILE(N2:N301, 0.025)

C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Samples											Median		Lower 95%	10.475
1	14	15	16	23	20	19	16	20	21	23	19.5		Upper 95%	19.7625
2	9	16	16	21	9	16	19	7	21	12	16			
3	14	14	9	16	21	21	17	16	7	14	15			
4	16	21	16	17	14	14	16	19	14	9	16			
5	12	14	10	10	14	9	21	23	19	19	14			
6	15	19	23	20	23	10	10	17	7	7	16			
7	21	15	14	23	21	10	7	16	15	20	15.5			
8	16	21	21	20	20	10	17	12	14	9	16.5			
9	16	12	20	17	17	7	15	21	10	14	15.5			
10	23	12	16	16	21	19	19	15	16	19	17.5			
11	19	12	12	10	21	20	21	10	10	19	15.5			
12	17	16	16	14	14	9	20	16	21	17	16			
13	9	9	19	7	21	10	14	23	20	9	12			
14	14	9	21	15	17	14	9	10	12	12	13			
15	14	21	10	21	21	10	19	17	14	9	15.5			
16	21	23	20	12	12	20	10	7	15	10	13.5			
17	12	17	23	10	9	15	19	16	21	14	15.5			
18	7	17	7	12	20	10	15	17	21	21	16			
19	12	16	23	16	16	16	15	16	21	10	16			
20	10	21	16	16	9	20	15	20	15	7	15.5			
21	15	21	16	9	7	14	16	16	21	14	15.5			

From the output we can see that the 95% bootstrapped confidence interval for the median value of the original dataset is .

Note that in this example we chose to generate 300 bootstrapped samples each with a sample size of n=10, but you can generate as many bootstrapped samples as you'd like.

When using statistical software, it's common to

generate thousands of bootstrapped samples which can then be used to construct a confidence interval.

Additional Resources

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