

Learn Bootstrapping Techniques in Excel: A Step-by-Step Guide

Authored by
Mohammed loot

October 29, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Learn Bootstrapping Techniques in Excel: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=5506>

Introduction to Bootstrapping in Statistics

Bootstrapping is a highly versatile and powerful statistical method designed to estimate the sampling distribution of a statistic. It achieves this by repeatedly resampling with replacement from the observed data. This technique becomes particularly invaluable when researchers face challenges such as small **sample sizes** or when the **underlying distribution** of the population is unknown, making traditional parametric statistical methods less reliable or even inapplicable. Its primary application lies in constructing **confidence intervals** and estimating other measures of accuracy for various statistical parameters.

At its core, the philosophy behind bootstrapping is to treat the observed dataset as a miniature representation of the entire population. By generating numerous new "resamples" from this original dataset, we effectively simulate the process of drawing multiple samples directly from the true, unknown population. This simulation is critical for understanding the inherent variability and uncertainty associated with our statistical estimates, even in the absence of complete population information.

Essentially, bootstrapping offers a non-parametric approach to statistical inference. This flexibility makes it an indispensable tool across diverse analytical contexts, especially when stringent assumptions about data distribution, often required by parametric tests, cannot be met. Its applications span various disciplines, from finance to environmental science, aiding in tasks like standard error estimation, confidence interval construction, and robust hypothesis testing.

The Fundamental Process of Bootstrapping Explained

The basic methodology of **bootstrapping** is built upon a sequence of clear, logical steps that can be applied to nearly any dataset. This process aims to empirically estimate the sampling distribution of a chosen **statistic**, which is essential for making robust statistical inferences. A thorough understanding of these foundational steps is paramount before moving into practical implementation.

The procedure initiates by considering your original, collected dataset as a surrogate for the entire population. From this empirical "population," a large number of new samples, known as bootstrap samples, are generated. Each of these bootstrap samples maintains the same size as the original dataset. A crucial aspect of this resampling technique is that it is always performed **with replacement**. This means that any single observation from your original dataset has the potential to be selected multiple times within a given bootstrap sample, and conversely, might not be selected at all for some samples.

Following the creation of these numerous bootstrap samples, the **statistic** of primary interest (e.g., the mean, **median**, standard deviation, or correlation coefficient) is calculated individually for each

of these bootstrap samples. This repetitive calculation yields a collection of estimates for that statistic, which collectively forms what is known as the bootstrap distribution. This generated distribution then serves as an empirical approximation of the true sampling distribution of the statistic, from which a [confidence interval](#) can be reliably derived.

Take k repeated samples [with replacement](#) from your given dataset.

For each of these k bootstrap samples, calculate the [statistic](#) you wish to estimate for the population.

This process generates k distinct estimates for your chosen statistic, forming a bootstrap distribution. This distribution is then used to compute a [confidence interval](#) for the statistic, providing a range of plausible values for the true population parameter.

The following step-by-step example will meticulously guide you through how to perform [bootstrapping](#) specifically within [Excel](#), illustrating each stage with clear, practical applications and explanations.

Step 1: Preparing Your Original Data in Excel

The foundational step for any statistical analysis, including [bootstrapping](#), is the accurate and organized input of your raw data. In the context of Excel, this means diligently entering your original dataset into a worksheet. This initial dataset is critically important as it will act as the master source from which all subsequent bootstrap samples are generated.

To begin, open a new Excel workbook. For optimal structure and ease of referencing in later steps, input your data into a single column, typically starting from cell A2. It is paramount to ensure that your data is clean, free from errors, and correctly formatted. Any inaccuracies at this stage could propagate through your analysis and compromise the validity of your bootstrap results.

For illustrative purposes in this example, we will utilize a hypothetical dataset. The image provided below demonstrates how this example dataset should be arranged and appear within your Excel spreadsheet. This specific layout is chosen to facilitate straightforward referencing by the Excel formulas that will be used to construct the bootstrap samples in the subsequent stages of the analysis.

	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: Generating Bootstrap Samples in Excel

Once your original data is meticulously entered and organized, the next pivotal stage is to generate a substantial number of [bootstrap samples](#). This process involves the random selection of values directly from your original dataset, critically employing the principle of [sampling with replacement](#). Leveraging Excel's powerful array of built-in functions enables us to automate this intricate, yet fundamental, resampling procedure.

To achieve this, we will combine the capabilities of three essential Excel functions: [INDEX](#), [RANDBETWEEN](#), and [ROWS](#). The formula provided below is specifically constructed to randomly pick a single value from your defined original data range.

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

Let's meticulously deconstruct this powerful formula to understand its components and functionality:

\$A\$2:\$A\$16: This segment precisely designates the range of cells that contains your original

dataset. The inclusion of absolute references (the dollar signs, \$) is absolutely vital. It ensures that this specified data range remains fixed and unchanging, even when you copy or drag the formula to other cells.

ROWS(\$A\$2:\$A\$16): This part of the formula dynamically computes the exact number of rows present within your designated data range. Its output serves as the crucial upper boundary for the random number generator, ensuring that only valid row indices are selected.

RANDBETWEEN(1, ROWS(\$A\$2:\$A\$16)): This function is responsible for generating a random integer. It produces a whole number between 1 (representing the first row of your data) and the total number of rows in your dataset (as determined by the ROWS function). This randomly generated integer effectively acts as a selector for a specific row within your data.

INDEX(range, row_num, column_num): Finally, the [INDEX function](#) takes center stage. It retrieves the specific value located at the intersection of the specified data range (\$A\$2:\$A\$16), the randomly determined row number (provided by RANDBETWEEN), and the first column (designated by 1, as our data is in a single column).

To initiate the generation of your samples, input this comprehensive formula into cell **D2**. This action will populate cell D2 with a value randomly selected from your original dataset. To construct your first complete bootstrapped sample (for instance, a sample of size 10), simply drag this formula horizontally to the right across 10 cells (from D2 to M2). Each cell in this row will now contain a distinct, randomly chosen value drawn from your original data.

To ensure a sufficient number of bootstrap samples for a robust statistical analysis, extend this process further. Drag this entire row of formulas (D2:M2) downwards. In the context of our example, we will drag it down through 300 rows, thereby creating 300 individual bootstrapped samples, each meticulously composed of 10 randomly selected values. This extensive resampling procedure forms the foundational bedrock of our bootstrap distribution.

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		Bootstrap 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	

It is absolutely imperative to reiterate a core principle: **bootstrapping** fundamentally relies on **sampling with replacement**. This methodological characteristic implies that a single data point from your original dataset can potentially appear multiple times within any given bootstrap sample. Conversely, it is also possible for certain values from the original dataset not to appear at all in some of the generated samples. This distinctive feature is intrinsic to the bootstrapping methodology, as it enables the effective simulation of the sampling distribution and captures the inherent variability of the data.

Step 3: Calculating the Statistic of Interest for Each Sample

With your extensive array of [bootstrap samples](#) now meticulously generated, the next crucial step involves calculating the specific [statistic](#) you are interested in for each individual sample. This collection of computed statistics will then coalesce to form your empirical bootstrap distribution, which is absolutely central to drawing meaningful inferences about the population parameter.

For the explicit purpose of this illustrative example, we will concentrate on computing the [median](#) value for each of the 300 bootstrap samples that we meticulously created in the preceding Step 2. The median is chosen here as a robust measure of central tendency; unlike the mean, it is less susceptible to distortion by extreme outliers, making it an excellent choice for demonstrating the practical application of the bootstrap process.

To execute this, navigate to the column immediately adjacent to your last bootstrap sample (for instance, if your samples span columns D through M, you would move to column N). In cell N2, you will enter the appropriate Excel [MEDIAN function](#), specifically referencing the first bootstrap sample. For example, your formula would look like `=MEDIAN(D2:M2)`. Once this formula is entered, drag it downwards for all 300 rows. This action will automatically calculate the median for every single one of your bootstrap samples, populating column N with these values.

N2															=MEDIAN(D2:M2)
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	Original Data		Bootstrap 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	

Upon the successful completion of this step, you will observe a new column (e.g., column N) that is entirely populated with the median values derived from each of your bootstrap samples. For instance, you might find:

The first bootstrapped sample yielding a median value of approximately **14**.

The second bootstrapped sample potentially resulting in a median value of about **16**.

The third bootstrapped sample perhaps producing a median value of **13.5**.

This comprehensive array of median values (or whatever other chosen [statistic](#)) constitutes your bootstrap distribution. It vividly represents the range of plausible values that your statistic could realistically assume, based on the inherent variability and characteristics observed within your original dataset.

Step 4: Constructing the Bootstrapped Confidence Interval

The culmination and arguably the most pivotal step in the [bootstrapping](#) methodology is the

construction of a [confidence interval](#) directly from the distribution of your meticulously calculated bootstrap statistics. A confidence interval serves as a probabilistic statement, providing a calculated range within which the true population parameter is highly likely to reside, coupled with a specified level of confidence.

When aiming to determine a 95% bootstrapped confidence interval, our objective is to identify the values that delineate the central 95% of our empirical bootstrap distribution. This translates into locating the 2.5th [percentile](#) and the 97.5th [percentile](#) of our collection of calculated statistics. These two specific percentile values will then establish the lower and upper boundaries of our desired confidence interval, respectively.

Excel's powerful [PERCENTILE function](#) is perfectly suited for this particular task. Assuming that your bootstrap medians are neatly arranged in column N, spanning from cell N2 down to N301, you would employ the following precise formulas to calculate the confidence interval bounds:

=PERCENTILE(N2:N301, 0.025)

=PERCENTILE(N2:N301, 0.975)

To obtain the lower bound of your 95% [confidence interval](#), input the first formula into an empty cell within your worksheet (for instance, cell P2). Subsequently, to determine the upper bound, enter the second formula into another empty cell (e.g., cell P3). These formulas will efficiently compute the values corresponding to the specified percentiles from your entire aggregation of 300 bootstrap medians, giving you the critical interval boundaries.

=PERCENTILE(N2:N301, 0.025)														
C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1 Samples											Median		Lower 95%	10.475
2	14	15	16	23	20	19	16	20	21	23	19.5		Upper 95%	19.7625
3	9	16	16	21	9	16	19	7	21	12	16			
4	14	14	9	16	21	21	17	16	7	14	15			
5	16	21	16	17	14	14	16	19	14	9	16			
6	12	14	10	10	14	9	21	23	19	19	14			
7	15	19	23	20	23	10	10	17	7	7	16			
8	21	15	14	23	21	10	7	16	15	20	15.5			
9	16	21	21	20	20	10	17	12	14	9	16.5			
10	16	12	20	17	17	7	15	21	10	14	15.5			
11	23	12	16	16	21	19	19	15	16	19	17.5			
12	19	12	12	10	21	20	21	10	10	19	15.5			
13	17	16	16	14	14	9	20	16	21	17	16			
14	9	9	19	7	21	10	14	23	20	9	12			
15	14	9	21	15	17	14	9	10	12	12	13			
16	14	21	10	21	21	10	19	17	14	9	15.5			
17	21	23	20	12	12	20	10	7	15	10	13.5			
18	12	17	23	10	9	15	19	16	21	14	15.5			
19	7	17	7	12	20	10	15	17	21	21	16			
20	12	16	23	16	16	16	15	16	21	10	16			
21	10	21	16	16	9	20	15	20	15	7	15.5			
22	15	21	16	9	7	14	16	16	21	14	15.5			

Based on the output generated by these Excel formulas, we can confidently determine the 95% bootstrapped [confidence interval](#) for the [median](#) value of the original dataset. In the specific context of this example, the estimated interval ranges from . This result implies that we can be 95% confident that the true population median lies somewhere within this calculated range.

It is important to acknowledge that for this practical demonstration, we opted to generate 300 bootstrapped samples, each consisting of a [sample size](#) of $n=10$. While 300 samples offer a clear illustration of the process, for more robust, statistically sound, and highly reliable results, it is a standard practice in professional statistical software to generate thousands, or even tens of thousands, of bootstrapped samples. The greater the number of samples you generate, the more accurately your bootstrap distribution will approximate the true sampling distribution, ultimately leading to more precise and trustworthy [confidence intervals](#).

Conclusion and Further Considerations

[Bootstrapping](#) stands out as a supremely flexible and powerful method for conducting statistical inference. It proves particularly advantageous in scenarios where the stringent assumptions of

traditional parametric methods cannot be met, or when researchers are contending with limited [sample sizes](#). As meticulously demonstrated through this guide, bootstrapping provides a practical, data-driven approach to constructing [confidence intervals](#) for a wide array of [statistics](#) directly from your observed data, entirely circumventing the need for prior knowledge about the [underlying distribution](#) of the population.

While Excel is an excellent tool for illustrating the fundamental principles and mechanics of bootstrapping, it is crucial to recognize that for more complex, large-scale analyses, and for generating the thousands of bootstrap samples often recommended for enhanced precision, specialized statistical software platforms are typically employed. Tools such as R, Python (with libraries like SciPy or scikit-learn), or SAS offer significantly more efficient computation, a broader spectrum of statistical functions, and advanced visualization capabilities, making them the preferred choice for professional-grade bootstrapping.

Regardless of the specific software or tool utilized, the core methodological steps of bootstrapping remain consistently robust: repeated resampling with replacement from the original dataset, calculating the statistic of interest for each generated resample, and subsequently deriving a robust inference (such as a confidence interval) from the resulting empirical bootstrap distribution. A thorough understanding and mastery of this technique will significantly enhance your capacity to perform resilient and reliable statistical analysis across various domains.

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