

Identifying the Top 5 Most Frequent Numbers in Excel: A Step-by-Step Guide

Authored by
Mohammed loot

November 9, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Identifying the Top 5 Most Frequent Numbers in Excel: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=14672>

Finding the single most frequent value (the [mode](#)) in a dataset using Microsoft Excel is straightforward. However, identifying the top N most frequent values--such as the **top 5**--requires a sophisticated combination of [Excel formulas](#), often involving an advanced technique known as an [Array Formula](#). This guide provides a detailed walkthrough of the necessary steps and explains the logic behind the powerful formula required to extract the top five most frequent numbers from any given range.

The core mechanism relies on iteratively finding the mode while excluding the values already identified as frequent. The master formula that accomplishes this complex task is structured as follows. Note that this formula must be entered as an array formula (using **Ctrl+Shift+Enter** in older Excel versions, though modern versions often handle it dynamically):

```
=MODE(IF(ISERROR(MATCH($A$2:$A$24,C$1:C1,0)),$A$2:$A$24))
```

When initially entered, this formula finds the single most frequently occurring value within the data range, specified here as **A2:A24**. By clicking and dragging this formula down into subsequent cells, we establish a dynamic mechanism to identify the second, third, fourth, and fifth most frequent numbers sequentially, effectively extracting the top 5 list. The absolute and relative cell references (e.g., **C\$1:C1**) are essential for this iterative process to function correctly.

The Challenge: Finding Multiple Modes in Excel

The standard [MODE function](#) in Excel is primarily designed to return only the single most frequently occurring number in a dataset. If a dataset exhibits multiple values that share the highest frequency (a multimodal distribution), or if we need to rank the top several frequencies, the basic MODE function is insufficient. To generate a ranked list of the top N most frequent items, we cannot simply repeat the basic MODE calculation; we must systematically exclude the results we have already found so that the function is forced to identify the next highest frequency among the remaining values.

This iterative exclusion is achieved by combining the [MODE function](#) with conditional logic functions--specifically, **IF**, [ISERROR](#), and [MATCH function](#)--to create an array that contains only the values that have not yet been identified in our results column. This sophisticated filtering mechanism ensures that each subsequent application of the formula searches the remaining pool of data points. By filtering out the previously identified modes, we guarantee that the formula successfully captures the next most frequent number down the statistical line, making this combination an indispensable tool for advanced frequency analysis.

Deconstructing the Advanced Array Formula

The complex formula is essentially a highly sophisticated filter built around the central statistical function, **MODE**. To truly master this technique, it is vital to understand the role of each nested component in excluding previously found values. The formula is `=MODE(IF(ISERROR(MATCH(A2:A24,C$1:C1,0)), A2:A24))`. The entire logic hinges on the **MATCH** and **ISERROR** components working together to dynamically shrink the dataset passed to **MODE**.

The [MATCH function](#) attempts to locate every single value in our main data range (**A2:A24**) within the results range built so far (**C\$1:C1**). When this formula is first placed in cell **C2**, the range **C\$1:C1** only refers to the cell immediately above, acting as an anchor. As the formula is dragged down, this reference expands (e.g., to **C\$1:C2**, then **C\$1:C3**, and so on), allowing the function to check if the current data point has already been identified as a top frequency number. If a match is found, the **MATCH** function returns a number (the position of the match); if no match is found, which indicates the value is new, it returns the **#N/A** error.

The [ISERROR function](#) then wraps the **MATCH** function. It converts the output of the **MATCH** function into a Boolean array: it returns **TRUE** if the value has NOT been found yet (because **MATCH** returned **#N/A**), or **FALSE** if the value HAS been found. This Boolean array is passed to the **IF** function, which acts as the filter. The **IF** statement uses the **TRUE/FALSE** results to determine which values from the original data range (**A2:A24**) should be passed to the [MODE function](#). Only values corresponding to **TRUE** (i.e., those not yet listed) are included in the array passed to **MODE**, while others are replaced by **FALSE**, effectively removing them from the mode calculation. This iterative filtering ensures that the result calculated is always the next highest frequency value not previously listed.

Step-by-Step Implementation: Setting Up the Data

To implement this robust method, we must first establish our dataset and the necessary result columns. Suppose we are working with the raw numerical data contained in Column A, specifically spanning the range **A2:A24**. Our goal is to extract the top five most frequent numbers into Column C. Proper setup, particularly the use of absolute references (the dollar signs) for the data range, is mandatory to ensure the formula remains stable when copied.

The following illustration provides a visual representation of the dataset used for this example. This data forms the foundational population for our frequency analysis. We must confirm that the data range specified in the formula, **\$A\$2:\$A\$24**, precisely matches the extent of our numerical list:

	A	B	C	D	E	F	G
1	Values						
2	4						
3	4						
4	5						
5	5						
6	15						
7	17						
8	7						
9	2						
10	15						
11	8						
12	4						
13	15						
14	17						
15	8						
16	9						
17	4						
18	3						
19	4						
20	15						
21	4						
22	16						
23	10						
24	5						
25							

A crucial detail for the iterative array formula is the management of the exclusion range starting at cell **C1**. If our results begin in cell **C2**, the formula's exclusion range starts by referencing **C\$1:C1**. Even if **C1** contains a header or is empty, it serves as the necessary, fixed starting point (**C\$1**) for the expanding relative reference (**C1**). This mechanism allows the range to correctly grow to **C\$1:C2**, **C\$1:C3**, and so on, as the formula is dragged down, ensuring that all previously identified frequent numbers are excluded from the current mode calculation. This meticulous setup is key to the success of the multi-mode extraction.

Calculating the Top 5 Most Frequent Numbers

The process of identifying the top frequencies is initiated by accurately entering the complete array formula into the first target cell, **C2**. This initial entry calculates the overall most frequent number in the entire range **A2:A24**. Depending on your version of Microsoft Excel, you must either confirm the entry by pressing **Ctrl+Shift+Enter** (which places curly braces around the formula, indicating it

is an [Array Formula](#)) or simply press Enter if using a newer version that supports dynamic arrays.

We input the full expression into cell **C2**, ensuring absolute and relative references are correctly placed:

=MODE(IF(ISERROR(MATCH(\$A\$2:\$A\$24,C\$1:C1,0)),\$A\$2:\$A\$24))

Once cell C2 correctly displays the highest frequency number, the subsequent results are easily generated. Simply click and drag the fill handle of cell C2 down to cell **C6**. Due to the intentional use of mixed references, Excel automatically adjusts the exclusion range in the formula for each subsequent row. Cell C3 calculates the mode after excluding the value in C2; C4 calculates the mode after excluding values in C2 and C3, and so on. This mechanism populates Column C with the five most frequent numbers in descending order of occurrence, providing the required ranked list:

	A	B	C	D	E	F	G
1	Values		5 Most Frequent Numbers				
2	4		4				
3	4		15				
4	5		5				
5	5		17				
6	15		8				
7	17						
8	7						
9	2						
10	15						
11	8						
12	4						
13	15						
14	17						
15	8						
16	9						
17	4						
18	3						
19	4						
20	15						
21	4						
22	16						
23	10						
24	5						
25							
26							

Validating Results: Determining the Frequency Count

Although Column C now contains the top five most frequent numbers, it lacks the critical context of how often each number actually appears in the dataset. To complete our analysis and validate the results, we must calculate the precise count for each identified number. This secondary step utilizes the highly efficient [COUNTIF function](#), which is perfectly suited for counting the number of cells within a specific range that meet a single specified criterion.

We will enter the COUNTIF formula into cell **D2**, directly adjacent to the first result in C2. The function requires two arguments: the range to be searched (the original data, which must be locked absolutely) and the criterion (the specific number to count, referenced relatively to Column C). The formula is constructed to efficiently look up the number listed in C2 within the static original dataset in Column A:

=COUNTIF(\$A\$2:\$A\$24, C2)

After entering this formula in D2, we use the autofill handle to drag it down to the remaining cells in column D (down to D6). Because the data range **\$A\$2:\$A\$24** is locked using absolute references, it remains constant across all cells. Conversely, the criterion reference (C2) updates automatically to C3, C4, C5, and C6, ensuring that the count performed in each row corresponds precisely to the unique number listed in Column C. This action populates Column D with the corresponding frequency count for each of the top 5 values.

D2		=COUNTIF(\$A\$2:\$A\$24, C2)			
	A	B	C	D	E
1	Values		5 Most Frequent Numbers	Frequency	
2	4		4	6	
3	4		15	4	
4	5		5	3	
5	5		17	2	
6	15		8	2	
7	17				
8	7				
9	2				
10	15				
11	8				
12	4				
13	15				
14	17				
15	8				
16	9				
17	4				
18	3				
19	4				
20	15				
21	4				
22	16				
23	10				
24	5				

Interpreting the Final Output

The resulting two-column table--listing the top frequency numbers in Column C and their verified counts in Column D--offers a complete and statistically meaningful summary of the most common data points within the original list. This final output is essential for data analysts who need more than just the single mode; they need to understand the concentration and ranking of values. By analyzing the data in this structured format, we can draw precise conclusions about the distribution of values in the dataset.

Based on the final results demonstrated in the image above, the hierarchy of frequencies is clearly defined:

The number **4** is confirmed as the primary mode of the dataset. It occurs **6 times**, making it the most dominant value in the list.

The number **15** is identified as the second most frequent value, appearing a total of **4 times** throughout the data range.

The number **5** holds the third position in frequency, having occurred **3 times**.

The remaining values listed in cells C5 and C6 represent the fourth and fifth most frequent numbers, providing a comprehensive ranking that reveals patterns of concentration often invisible to standard statistical functions.

This methodology successfully leverages a complex [Array Formula](#) structure to move beyond simple mode calculation and extract a ranked list of the top N frequencies. This technique is invaluable for advanced data analysis in Excel when standard functions fall short of providing deep insight into data distribution.

Additional Resources for Frequency Analysis

Mastering frequency counting and conditional statistical calculations, like those demonstrated here, is fundamental for users aiming for advanced data proficiency in Excel. The following resources offer additional information and alternative methods on how to count specific criteria and analyze frequencies within large datasets, enabling you to apply these skills to even more complex data management challenges.