

Learn How to Calculate Averages in Excel While Excluding Outliers

Authored by
Mohammed looti

October 29, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Learn How to Calculate Averages in Excel While Excluding Outliers*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=5709>

Introduction: Understanding Outliers and Their Impact on Averages

When conducting in-depth analysis of any [dataset](#), analysts frequently encounter the challenge posed by statistical [outliers](#). These are defined as data points that deviate significantly from the majority of other observations within the distribution. An outlier can dramatically skew common statistical measures, such as the arithmetic [average](#) or the [mean](#), potentially leading to inaccurate conclusions regarding the central tendency of the underlying process. Such extreme values may stem from genuine variability, measurement error, or recording mistakes. Regardless of the root cause, failing to manage outliers can lead to misleading interpretations, making their appropriate handling a critical step in accurate data reporting.

To obtain a more representative measure of central tendency—one that accurately reflects the typical value of the core data—it is often necessary to systematically exclude these extreme values before calculation. [Microsoft Excel](#), recognized globally as a principal tool for data manipulation and analysis, provides several powerful, built-in functions specifically designed to address this requirement. This comprehensive guide will meticulously explore two distinct and highly effective statistical approaches available in Excel for calculating an average while robustly mitigating the influence of these distorting data points.

These two primary methods leverage different statistical methodologies, offering flexibility based on the dataset's characteristics and the specific analytical objectives. We will demonstrate:

The use of the **TRIMMEAN** function, which automatically calculates a [trimmed mean](#) by excluding a user-defined, fixed percentage of observations from both the upper and lower extremes of the dataset.

The application of the statistically rigorous [Interquartile Range \(IQR\)](#) method, which first identifies outliers based on their deviation from the central 50% of the data, and then excludes only those specific points from the subsequent average calculation.

Throughout this tutorial, we will utilize a single, consistent sample dataset to ensure a clear, side-by-side comparison of the results generated by each method. Understanding how these techniques operate and how they impact the final calculated average is essential for anyone responsible for producing reliable statistical reports and informed data interpretations.

Preparing the Sample Dataset for Analysis

To provide a practical and easily replicable foundation for this tutorial, we will work with a simple numerical dataset structured within a standard Excel worksheet. This dataset comprises 15 individual values, which will serve as the common input for both the TRIMMEAN and the IQR-based averaging methods. Before implementing any complex formulas, it is beneficial to visually inspect the data to understand the underlying spread and identify potential candidates for

exclusion.

	A	B	C	D	E
1	data				
2	18				
3	24				
4	26				
5	34				
6	38				
7	45				
8	48				
9	54				
10	60				
11	73				
12	79				
13	85				
14	94				
15	98				
16	164				
17					
18					
19					
20					
21					

Upon observation, it is clear that these values span a range, with some observations appearing notably higher or lower than the bulk of the data points. For instance, values like 164 or 12 are likely to exert a disproportionate gravitational pull on a standard arithmetic mean. Our overarching analytical objective is to calculate an average that is significantly less susceptible to the influence of these potentially extreme values, thereby achieving a more stable, representative, and unbiased measure of the typical value present in the core data distribution.

Method 1: The Robustness of the TRIMMEAN Function

The first and simplest technique for outlier exclusion involves leveraging the [TRIMMEAN](#) function, a standard statistical tool available in Excel. This function is specifically designed to compute the [trimmed mean](#), which is the [mean](#) of a dataset after a specified percentage of observations have been removed from both the high and low ends of the sorted distribution. This method offers inherent robustness, making it highly valuable when analysts suspect that extreme values may be disproportionately influencing the simple arithmetic average.

The functional syntax for TRIMMEAN is concise: `TRIMMEAN(array, percent)`. The `array` parameter refers to the range of numerical values intended for averaging, while the `percent` parameter specifies the total fractional proportion of data points to be excluded from the calculation. For example, if an analyst specifies 20% for the `percent` argument, the function will remove 10% of the observations from the lower tail and 10% from the upper tail of the sorted dataset. This approach is widely favored due to its simplicity and its ability to provide a consistent, more reliable average without requiring the complex, explicit identification of individual outliers based on statistical rules.

We will now apply this function to our sample dataset, contained in column A (A2:A16). We choose to calculate the average while excluding a cumulative total of 20% of the observations. This decision results in the removal of 10% from the smallest values and 10% from the largest values. The precise formula entered into Excel is:

=TRIMMEAN(A2:A16, 20%)

Given that our dataset comprises 15 values, calculating 20% of 15 yields 3 observations in total (1.5 observations at each end). The TRIMMEAN function is engineered to round the number of excluded observations down to the nearest integer. Consequently, one value (1) is removed from the bottom and one value (1) is removed from the top of the sorted list, leaving 13 core values remaining for the average computation. By automatically trimming the extremes, this function successfully mitigates the distorting effect of the smallest and largest data points, thus yielding a centralized average.

C2				=TRIMMEAN(A2:A16, 20%)		
	A	B	C	D	E	F
1	data		Avg. Excluding Outliers			
2	18		58.30769			
3	24					
4	26					
5	34					
6	38					
7	45					
8	48					
9	54					
10	60					
11	73					
12	79					
13	85					
14	94					
15	98					
16	164					
17						
18						
19						
20						

The outcome of this calculation demonstrates that the average, after excluding the potential outliers using the robust TRIMMEAN function, is precisely **58.30769**. This figure offers a substantially more stable and representative measure of the central tendency when compared to a simple, untrimmed average, particularly in datasets known to contain influential extreme values that would otherwise inflate or deflate the mean calculation.

Method 2: Identifying Outliers Using the Interquartile Range (IQR)

The second methodology offers a statistically precise, data-driven mechanism for outlier removal by utilizing the [Interquartile Range \(IQR\)](#). The IQR serves as a crucial measure of statistical dispersion, quantifying the spread of the middle 50% of a [dataset](#). It is calculated as the difference between the third [quartile](#) (Q3, the 75th [percentile](#)) and the first quartile (Q1, the 25th percentile). Because it focuses solely on the central distribution, the IQR provides a measure of variability that is notably less sensitive to the influence of extreme values than metrics such as the standard deviation.

The standard statistical rule for defining an observation as an outlier--often referred to as the Tukey method--states that any data point is considered extreme if it falls below the lower fence ($Q1 - 1.5 *$

IQR) or rises above the upper fence ($Q3 + 1.5 * IQR$). These boundaries provide a mathematically sound basis for outlier detection, allowing the analyst to exclude only those specific points that statistically deviate beyond a reasonable threshold. This approach is fundamentally different from TRIMMEAN, as it adapts dynamically to the data's actual distribution rather than relying on a fixed removal percentage.

The initial step in implementing the IQR method in Excel requires calculating the Q1, Q3, and the IQR itself. We utilize the QUARTILE.INC function for accuracy and consistency across modern Excel versions. The IQR is derived by simply subtracting the value of the first quartile from the value of the third quartile, using the following formula referencing our data range A2:A16:

=QUARTILE(A2:A16,3)-QUARTILE(A2:A16,1)

Once computed, this formula yields the numeric difference between the 75th and 25th percentile values, establishing the core measure of data spread. This resulting IQR value is absolutely essential for the subsequent calculation of the outlier fences, which determine the precise thresholds for identifying extreme data points.

	A	B	C	D	E	F	G
1	data						
2	18						
3	24						
4	26						
5	34						
6	38						
7	45						
8	48						
9	54						
10	60						
11	73						
12	79						
13	85						
14	94						
15	98						
16	164						
17							
18	IQR	46	=QUARTILE(A2:A16, 3)-QUARTILE(A2:A16, 1)				
19							
20							
21							
22							

With the IQR successfully determined, the next complex step involves creating a column that flags

each individual data point as either an outlier (1) or a non-outlier (0). This requires using an intricate conditional statement that checks if the data point violates either the upper or the lower fence criteria. We employ a combination of the IF and OR logical functions, ensuring that we use absolute references (e.g., \$A\$2:\$A\$16 and \$B\$18 for the IQR cell) so that the formula can be accurately copied down the column:

=IF(OR(A2<QUARTILE(\$A\$2:\$A\$16,1)-1.5*\$B\$18,A2>QUARTILE(\$A\$2:\$A\$16,3)+1.5*\$B\$18),1,0)

When this formula is implemented adjacent to the dataset, it systematically evaluates whether the value in cell A2 is less than the calculated lower fence OR greater than the calculated upper fence. If either condition is true, the cell receives a "1"; otherwise, it receives a "0". This process provides a clear, binary indicator of outlier status across the entire dataset.

	A	B	C	D	E	F	G	H	I	J
1	data	outlier?								
2	18	=IF(OR(A2<QUARTILE(\$A\$2:\$A\$16,1)-1.5*\$B\$18,A2>QUARTILE(\$A\$2:\$A\$16,3)+1.5*\$B\$18),1,0)								
3	24	0								
4	26	0								
5	34	0								
6	38	0								
7	45	0								
8	48	0								
9	54	0								
10	60	0								
11	73	0								
12	79	0								
13	85	0								
14	94	0								
15	98	0								
16	164	1								
17										
18	IQR	46								
19										
20										
21										
22										
23										

The application of this rigorous statistical test reveals that, within our specific sample, only one value--the observation of **164**--is definitively identified as an outlier. This meticulous identification process is the chief advantage of the IQR method, as it allows us to target and remove only those data points that statistically warrant exclusion, contrasting sharply with the blunt, fixed-percentage removal characteristic of the TRIMMEAN function.

Calculating the Refined Average After IQR Exclusion

Following the successful identification of the specific outlier(s) using the robust IQR methodology, the concluding step involves calculating the average of the remaining data points--those assigned a "0" indicator, signifying they are statistically within the normal range. For this selective averaging task, Excel's powerful [AVERAGEIF](#) function is the ideal tool, as it permits the calculation of an average based exclusively on a specified, logical criterion.

We configure the AVERAGEIF function to average all numerical values located in our original data range (column A) only when the corresponding cell in the adjacent outlier indicator column (column B) contains the criterion value of "0." This meticulous filtering ensures that the calculation is performed exclusively on the statistically "normal" data points, resulting in a highly refined and accurate measure of central tendency that is completely free from the distorting influence of the identified extreme values. The formula implemented is as follows:

=AVERAGEIF(B2:B16, 0, A2:A16)

In this construction, B2:B16 serves as the criteria range (the outlier indicators), 0 is the specific criterion (to include only non-outliers), and A2:A16 is the actual range containing the numerical values to be averaged. This selective application ensures the calculated [mean](#) is truly representative of the core data distribution.

C2			$\times$	$\checkmark$	fx	<code>=AVERAGEIF(B2:B16, 0, A2:A16)</code>	
	A	B	C	D	E	F	G
1	data	outlier?	Avg. Excluding Outliers				
2	18	0	55.42857				
3	24	0					
4	26	0					
5	34	0					
6	38	0					
7	45	0					
8	48	0					
9	54	0					
10	60	0					
11	73	0					
12	79	0					
13	85	0					
14	94	0					
15	98	0					
16	164	1					
17							
18	IQR	46					
19							
20							
21							
22							
23							

The final average of the dataset, meticulously calculated after the IQR-based exclusion of the identified outlier, is determined to be **55.42857**. This result clearly diverges from the outcome produced by the TRIMMEAN function (58.30769), illustrating the fundamental methodological differences between the two techniques. This comparison underscores the necessity of selecting the most appropriate outlier handling method based on the analytical goals and the desired level of statistical justification.

Comparing Methodologies and Determining Best Practices

Both the TRIMMEAN function and the Interquartile Range (IQR) method offer statistically valid mechanisms for computing an average while simultaneously neutralizing the impact of extreme values. However, they rely on fundamentally distinct principles, making each method ideally suited for different analytical contexts. Understanding these operational differences is paramount for any analyst aiming to select the most appropriate technique for their specific [dataset](#).

TRIMMEAN: This method excels in terms of simplicity and expediency. It provides a quick, inherently robust average by applying a fixed percentage exclusion to both tails of the data

distribution. This is advantageous when a rapid calculation is needed and the analyst is comfortable with removing a predetermined quantity of data, regardless of whether those points are statistically true outliers or merely part of the natural data variability. However, its primary limitation is this very fixed nature; it can inadvertently discard valid observations if the distribution is skewed but does not contain extreme outliers. TRIMMEAN is best utilized when a consistently conservative average is required across large datasets where manual outlier investigation is impractical.

Interquartile Range (IQR): The IQR method provides a statistically rigorous and highly justifiable procedure for outlier detection. It identifies extreme values based on a clear, universally accepted statistical rule (the $1.5 * \text{IQR}$ criterion). This ensures that only values that are objectively distant from the central 50% of the data are flagged and subsequently excluded. Although the implementation of the IQR method in Excel is more complex, requiring multiple steps, the payoff is precision: the resulting average is based on a refined dataset where exclusion is statistically warranted. This method is the preferred choice when precision in outlier identification is critical, and the analysis requires strong statistical justification for the removal of observations.

When selecting between these two powerful methods, analysts must consider both the volume and nature of the data, as well as the ultimate goal of the analysis. If the objective is rapid visualization or general robustness, TRIMMEAN offers efficiency. If the analysis demands precise statistical validity, such as in academic research or formal reporting where justifying the removal of data is mandatory, the IQR method is overwhelmingly the superior choice. Crucially, analysts must always maintain a deep understanding of their data's context and the implications of any observation exclusions before drawing final conclusions.

Advancing Your Expertise in Excel Statistics

The ability to accurately handle and exclude outliers is a fundamental skill, yet it represents only one facet of advanced statistical analysis within Excel. The techniques explored here provide a strong foundation for moving toward more sophisticated data computations and modeling. To further enhance your analytical capabilities and ensure you extract the most accurate insights, it is highly recommended to continue exploring other common data manipulation and statistical tasks within Excel.

Expanding your proficiency in related statistical functions will empower you to perform deeper diagnostics and validation on your datasets. Consider delving into the following related topics to broaden your Excel skillset:

Understanding Skewness and Kurtosis in Excel
Calculating Moving Averages in Excel
Advanced Data Validation Techniques in Excel

Continuous investment in mastering these advanced Excel capabilities will significantly bolster your capacity to perform complex data analysis, validate findings, and ultimately make more informed, data-driven decisions.