

# Learn to Calculate Averages in Google Sheets by Excluding Outliers

Authored by  
**Mohammed loot**

February 22, 2026

## RECOMMENDED CITATION

Mohammed loot (2026). *Learn to Calculate Averages in Google Sheets by Excluding Outliers*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=3108>

## Introduction: Mastering Outlier Exclusion in Google Sheets

In the crucial realm of data analysis, achieving accurate and truly representative statistical measures is absolutely paramount. However, real-world datasets are frequently contaminated by **outliers**--data points that stand significantly apart from the general body of observations. These extreme values possess the power to severely skew critical calculations, most notably the **average**, often resulting in conclusions that are misleading or unreflective of the core phenomenon being studied. Fortunately, sophisticated spreadsheet platforms like **Google Sheets** offer robust, built-in functionalities designed specifically to identify, handle, and effectively exclude such statistical anomalies, thereby ensuring that your resulting analyses are reliable and accurately reflect the underlying data distribution.

This comprehensive guide is dedicated to exploring two distinct yet equally powerful methodologies for calculating a central tendency measure, or average, while systematically excluding these detrimental outliers directly within the Google Sheets environment. Each method utilizes a unique analytical perspective, catering to different levels of statistical rigor and varying data characteristics. By thoroughly understanding both techniques, analysts will be equipped with a versatile toolkit for achieving cleaner, more meaningful data interpretation and decision-making.

Specifically, we will provide step-by-step instructions on how to leverage these two strategies:

Utilizing the **TRIMMEAN function**, which acts as a straightforward statistical shortcut designed to automatically trim a specified, arbitrary percentage of data points from both the upper and lower extremes of your dataset.

Employing the **Interquartile Range (IQR)** method, which represents a more statistically rigorous and data-driven approach that identifies genuine outliers based on their measured distance from the central tendency of the data distribution.

To ensure a practical illustration of these advanced statistical techniques, we will consistently apply both methods to the same, standardized dataset provided below. This sample data serves as our common operational ground for demonstrating the precise application, step-by-step execution, and comparative results of both outlier exclusion strategies in Google Sheets.

	A	B	C	D	
1	<b>data</b>				
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					

## Understanding Outliers and Their Impact on Central Tendency

Before proceeding to the technical implementation of outlier exclusion formulas, it is crucial to establish a clear conceptual understanding of what outliers are and why they pose a significant challenge to standard data analysis. Fundamentally, an **outlier** is an observation point that exhibits extreme deviation or distance from other observations in a sample. These unusual points can originate from diverse sources, including simple measurement errors, mistakes during data entry, or, occasionally, genuine and meaningful, albeit extreme, variations inherent within the population under observation. While some outliers may hold valuable, unique insights, others can severely destabilize and distort standard statistical calculations.

The mere presence of one or two outliers can disproportionately impact descriptive statistics, particularly the calculated **arithmetic mean**. A single extremely high or extremely low value acts as a statistical magnet, pulling the average in its direction and rendering it unrepresentative of the vast majority of the data points. Consider, for instance, calculating the average net worth within a specific community: the inclusion of a single multi-billionaire's assets in the calculation would drastically inflate the perceived average wealth for most residents, leading to a highly inaccurate

representation of the typical financial standing within that neighborhood.

Therefore, proactively identifying and appropriately handling these influential data points is a non-negotiable step in maintaining the integrity and accuracy of your data analysis. By systematically excluding these extreme values when deriving measures of central tendency, analysts can calculate a more robust and meaningful average that accurately reflects the typical characteristics and patterns of the dataset. The following sections provide detailed guidance on the practical application of the two primary methods used to achieve this critical data cleaning step in Google Sheets.

## Method 1: Utilizing the TRIMMEAN Function for Robust Averages

The first and simplest technique we will explore for calculating an average while mitigating the influence of outliers involves the specialized [TRIMMEAN function](#) available in Google Sheets. This function is an efficient statistical tool designed to compute the mean of a dataset only after a specified percentage of observations has been automatically removed from both the upper and lower tails of the data distribution. It is especially useful in situations where analysts suspect the presence of extreme values but prefer a swift, generalized solution rather than needing to identify specific outliers individually.

The syntax and operation of the [TRIMMEAN function](#) are straightforward, requiring only two arguments: the range of numerical values intended for analysis, and the percentage of data (expressed as a decimal or percentage) that should be collectively excluded. This specified percentage is then divided equally and removed from the smallest values (the lower tail) and the largest values (the upper tail) within the dataset. For example, if an analyst specifies a 20% trimming percentage, the function will remove 10% of the observations from the bottom and 10% from the top of the sorted data, effectively neutralizing the undue influence of extreme values on the resulting average.

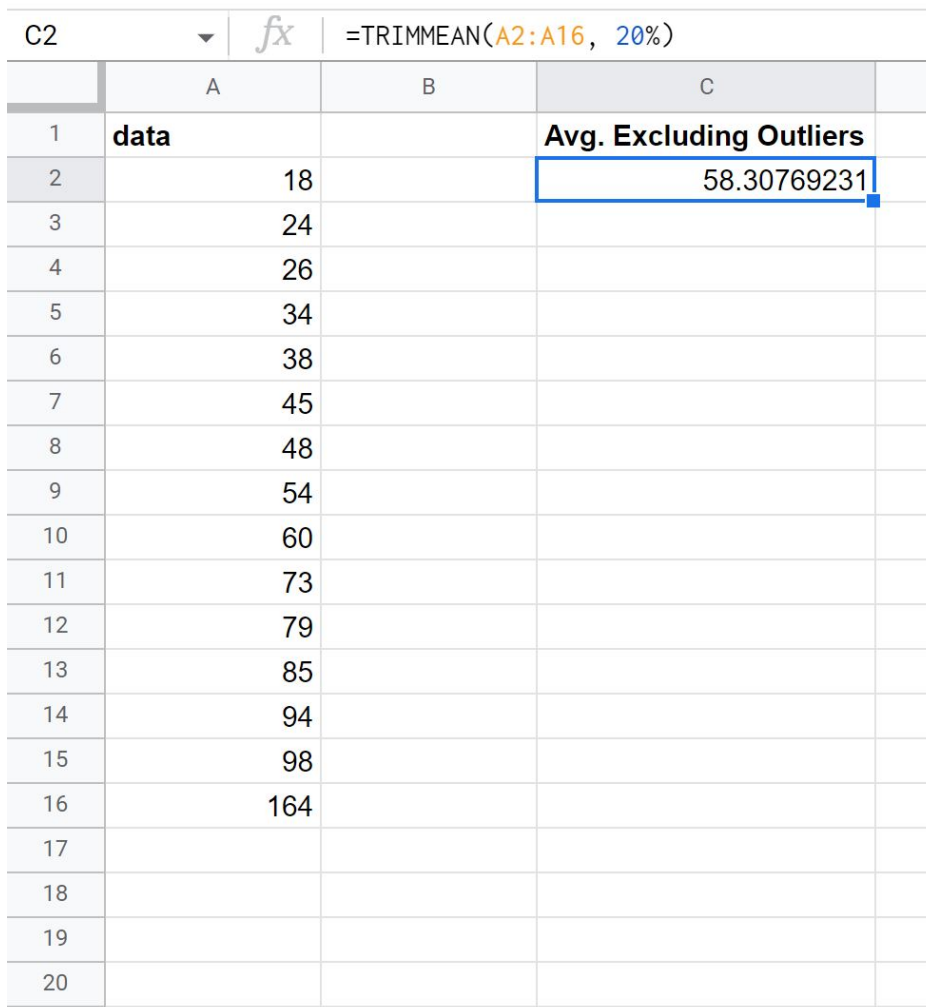
Let us apply this function to our provided sample dataset, which resides in column A. To calculate the average value while excluding a cumulative total of 20% of the observations (meaning 10% from the highest values and 10% from the lowest), we would construct and utilize the following precise formula:

**=TRIMMEAN(A2:A16, 20%)**

It is important for the analyst to clearly understand how the trimming percentage translates into the actual number of observations that are removed. Our specific dataset contains 15 total values. When we specify 20% for trimming, the function calculates 20% of 15, which results in 3 total observations to be removed. Since the trimming must be applied equally to both ends, 1.5

observations would technically be removed from the top and 1.5 from the bottom. Google Sheets' [TRIMMEAN function](#) is designed to round this calculated number down to the nearest whole number for each side, meaning it will exclude the single smallest value and the single largest value (1 observation from each end, totaling 2 observations removed, as 1.5 rounds down to 1). This ensures that the function operates predictably with discrete data points.

Once this formula is executed, Google Sheets efficiently processes the dataset, removes the specified count of extreme values, and then computes the resulting average of the remaining data points. The screenshot provided below clearly illustrates this calculation and the resulting outcome:



	A	B	C
1	data		<b>Avg. Excluding Outliers</b>
2	18		58.30769231
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			

As clearly depicted in the figure, the calculated average with outliers excluded using the [TRIMMEAN function](#) yields a value of **58.30769**. This statistically trimmed mean provides a significantly more robust measure of central tendency compared to a simple, unadjusted arithmetic mean, as its calculation is inherently less susceptible to the distorting influence of potential extreme values residing at the fringes of the dataset.

## Method 2: Employing the Interquartile Range (IQR) for Precise Outlier Detection

The second methodology provides a statistically superior and data-driven framework for the precise identification and exclusion of outliers, accomplished through the calculation and utilization of the [Interquartile Range \(IQR\)](#). The [IQR](#) serves as a key measure of statistical dispersion, specifically representing the range that encompasses the middle 50% of the data distribution. Mathematically, it is derived by calculating the difference between the third [quartile \(Q3\)](#) and the first [quartile \(Q1\)](#) within a given dataset.

The [IQR](#) is exceptionally valuable for outlier detection because, unlike the standard deviation, it is highly resistant to the influence of extreme values. The universally accepted statistical rule for identifying outliers, typically attributed to statistician John Tukey, dictates that an observation is classified as an outlier if it falls below the calculated lower fence ( $Q1 - 1.5 \times IQR$ ) or above the calculated upper fence ( $Q3 + 1.5 \times IQR$ ). This established threshold method helps define objective boundaries beyond which data points are considered significantly unusual relative to the central bulk of the data.

To implement this method, the initial requirement is to calculate the [IQR](#) for our dataset within Google Sheets. The dedicated `QUARTILE` function is essential for this step: `QUARTILE(data, 1)` retrieves Q1 (the 25th percentile), and `QUARTILE(data, 3)` retrieves Q3 (the 75th percentile). Consequently, the formula required to calculate the [IQR](#) is:

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

The following screenshot demonstrates the precise implementation of this formula in your Google Sheet, which calculates and displays the resulting [IQR](#) value:

B18		fx =QUARTILE(A2:A16,3)-QUARTILE(A2:A16,1)			
	A	B	C	D	
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		
19					
20					
21					
22					
23					

Once the **IQR** value is accurately determined, the subsequent critical step is to precisely identify which specific values in our dataset meet the criteria to be classified as outliers based on the rigorous  $1.5 \times \text{IQR}$  rule. This requires constructing a complex **conditional formula** that tests each individual data point against both the calculated lower and upper fences. For ease of subsequent averaging, we will assign a flagging value of "1" to any value that is confirmed as an outlier and "0" to those that are not. This efficient flagging mechanism systematically prepares the data for the final averaging calculation.

The formula provided below strategically utilizes the `IF` and `OR` logical functions to execute this critical check. It meticulously compares each value in column A (e.g., cell `A2`) against the derived lower bound ( $Q1 - 1.5 \times \text{IQR}$ ) and the upper bound ( $Q3 + 1.5 \times \text{IQR}$ ). It is essential to ensure the use of absolute references (e.g., `\$A\$2:\$A\$16`, `\$B\$18`) for the data range and the cell containing the calculated **IQR** value, guaranteeing that these references remain static when the formula is dragged down across the entire 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)
```

The subsequent screenshot illustrates the successful application of this formula across the entire dataset. A new column is typically allocated to house these generated flags, providing a transparent and immediate indication of which specific observations are deemed statistically significant outliers:

	A	B	C	D	E
B2		=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)			
1	<b>data</b>	<b>outlier?</b>			
2	18	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	<b>IQR</b>	46			
19					
20					
21					
22					

Following the application of the conditional formula, we can clearly observe that, within our specific dataset, the only single value that qualifies as an outlier according to the IQR rule is **164**. This precise and statistically justified identification allows us to exclude only the genuinely extreme value without resorting to the arbitrary removal of a percentage of the overall data.

Finally, with the outliers definitively flagged, we can proceed to calculate the average of all values in the dataset that have been confirmed as non-outliers. The [AVERAGEIF function](#) is perfectly optimized for this exact task, as it enables the averaging of values based on a specific, predetermined criterion--in this scenario, only including data points where our outlier flag is exactly "0".

The final formula required for calculating this statistically refined average is as follows:

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

In this formula structure, the range `B2:B16` specifies the cells containing our outlier flags (1 for outlier, 0 for not), the criterion `0` restricts the calculation to include only non-outliers, and the range `A2:A16` is the actual set of numerical values we wish to average.

The concluding screenshot demonstrates the application of this final formula, successfully yielding the average of the dataset with the single identified outlier fully excluded:

C2				
fx				
=AVERAGEIF(B2:B16, 0, A2:A16)				
	A	B	C	D
1	<b>data</b>	<b>outlier?</b>	<b>Avg. Excluding Outliers</b>	
2	18	0	55.42857143	
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	<b>IQR</b>	46		
19				
20				
21				

As a result of this meticulous process, the average with outliers excluded using the [Interquartile Range](#) method is calculated as **55.42857**. This derived value represents a significantly more accurate [average](#) of the typical data points, now free from the distorting influence of the single, identified extreme value.

## Comparing the Two Methods and Best Practices

Having thoroughly investigated both the [TRIMMEAN function](#) and the [Interquartile Range \(IQR\)](#) methods for systematically excluding outliers in Google Sheets, it is imperative to analyze their respective operational strengths and inherent weaknesses to determine which approach is most suitable for different analytical scenarios. While both methods successfully aim to provide a more robust [average](#), they fundamentally achieve this goal through diverging statistical principles.

The [TRIMMEAN function](#) excels in its notable simplicity and efficiency of implementation. It requires only a single, compact formula, making it the preferred choice for rapidly mitigating the general impact of extreme values. However, its primary analytical limitation lies in the subjective nature of the trimming percentage. The decision regarding what percentage to trim (e.g., 10%, 20%) is often arbitrary and is not directly derived from or justified by the intrinsic statistical properties of the data itself. If the selected percentage is too conservatively low, genuine outliers may remain influential; conversely, if the percentage is excessively high, valuable data points that are not true anomalies might be inadvertently excluded, potentially masking genuine underlying variability or important observations.

In contrast, the [IQR](#) method is inherently more statistically rigorous, providing an objective, data-driven criterion for [outlier](#) identification. By strictly adhering to the established  $1.5 \times \text{IQR}$  rule, this method offers a highly objective and precise mechanism to determine which points are truly anomalous. This approach provides greater precision by exclusively targeting extreme values, thus preserving a larger proportion of the original, non-outlying data. The main drawback of the IQR method is its increased complexity, which involves multiple necessary steps and formulas (calculating Q1/Q3, determining IQR, setting up conditional fences, flagging outliers, and finally averaging), demanding a deeper foundational understanding of statistical concepts and a more complex setup within Google Sheets.

The optimal choice between these two powerful methods must ultimately align with your specific analytical objectives and the inherent characteristics of your dataset. If the requirement is a quick, generalized solution to reduce the influence of extremes, and you are comfortable with a percentage-based trimming approach, [TRIMMEAN](#) is undoubtedly an excellent choice. If, however, precision, statistical rigor, and the ability to definitively identify and isolate only genuine outliers are paramount, the [IQR](#) method is strongly preferable, despite the associated increase in implementation complexity. Regardless of the chosen technique, a standard best practice involves visually inspecting your data (e.g., through the use of box plots or scatter plots) to gain supplementary insights into the overall data distribution and the specific nature of any observed anomalies.

## Conclusion and Further Exploration

The capacity to effectively handle and mitigate the influence of **outliers** is an absolutely critical skill for any professional working with quantitative data, and Google Sheets provides highly capable tools to accomplish this essential task. We have systematically explored two distinct yet highly effective methods for robust average calculation: the operationally straightforward **TRIMMEAN function** and the statistically nuanced **Interquartile Range (IQR)** approach. Each technique offers a reliable pathway to calculate an **average** that is significantly less susceptible to distortion from extreme data points, thereby enhancing the overall reliability and trustworthiness of your analytical insights.

By achieving a solid understanding of both the simplicity inherent in trimming a fixed percentage of data and the statistical precision involved in identifying outliers based on their calculated distance from the quartiles, you are now better equipped to make informed and methodologically sound decisions about data cleansing. The final selection between these two powerful methods should often hinge on the specific context of your ongoing analysis, the desired level of statistical rigor required, and the interpretability demanded of the final results. We strongly encourage readers to immediately apply these demonstrated techniques to their own datasets and observe firsthand how these robust methods can refine their understanding of underlying trends and central tendencies.

For those interested in further expanding their Google Sheets proficiency and exploring other common data manipulation and analytical tasks, the following tutorials provide additional valuable resources: