

How to Identify Outliers in Excel Using Grubbs' Test

Authored by
stats writer

March 5, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Identify Outliers in Excel Using Grubbs' Test*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=134097>

How can I conduct Grubbs' Test in Excel?

The **Grubbs' Test**, also widely recognized as the maximum normed residual test, serves as a fundamental **statistical test** utilized to detect a single **outlier** within a univariate dataset that follows an approximately **normal distribution**. In the realms of quality control, laboratory testing, and general data science, identifying these anomalies is crucial because a single extreme value can significantly distort the **sample mean** and inflate the **standard deviation**, leading to erroneous conclusions. By employing this methodology, researchers can mathematically determine if a suspicious data point is truly a statistical anomaly or merely a natural variation within the expected range of the population.

To successfully implement the **Grubbs' Test**, the dataset under investigation should ideally consist of at least seven observations to ensure the results possess sufficient **statistical power**. It is important to note that this specific procedure is designed to identify only one **outlier** at a time. If you suspect that your data contains multiple anomalous points, you should consider alternative methods such as the **Tietjen-Moore test** or the **Generalized Extreme Studentized Deviate (ESD) test**, as the standard **Grubbs' Test** may suffer from "masking," where the presence of multiple outliers makes each one appear less significant than it actually is.

When conducting this analysis in **Microsoft Excel**, the process involves comparing a calculated test statistic against a **critical value** derived from the **Student's t-distribution**. This comparison allows us to either reject or fail to reject the **null hypothesis**, which states that there are no outliers in the dataset. By automating these calculations within a spreadsheet, data analysts can create reusable templates that streamline the data cleaning process across various projects, ensuring higher **data integrity** and more reliable statistical outcomes.

The Theoretical Framework of the Grubbs' Test Statistic

The calculation of the **Grubbs' Test** statistic, denoted as **G**, depends entirely on which end of the distribution the suspected **outlier** resides. If the researcher suspects that the maximum value in the dataset is the anomaly, the test statistic is determined by calculating the difference between the maximum value and the **sample mean**, then dividing that result by the **sample standard deviation**. This formula essentially quantifies how many standard deviations the maximum point lies away from the average, providing a standardized measure of its extremity.

Conversely, if the minimum value is under suspicion, the formula is adjusted to subtract the minimum value from the **sample mean** before dividing by the **standard deviation**. In scenarios where the researcher does not have a prior hypothesis about which direction the **outlier** might appear, a two-sided test is appropriate. In this case, the test statistic **G** is defined as the maximum absolute deviation from the mean divided by the **standard deviation**. This versatile approach

ensures that the most extreme value in either direction is scrutinized with equal rigor.

Mathematically, these calculations are represented as follows: For a suspected maximum **outlier**, $G = (x_{\max} - \bar{x}) / s$. For a suspected minimum **outlier**, $G = (\bar{x} - x_{\min}) / s$. For a two-tailed evaluation, $G = \max|x_i - \bar{x}| / s$. Here, \bar{x} represents the **sample mean** and s represents the **standard deviation**. Understanding these variations is essential for applying the correct logic within **Excel** to ensure the **statistical significance** of your findings is accurately captured.

Establishing the Critical Value for Statistical Comparison

Once the test statistic G is calculated, it must be compared to a **critical value** to determine if the result is statistically significant. The formula for the **critical value** of the **Grubbs' Test** is somewhat complex, as it incorporates the sample size and a specific **significance level**. The calculation is as follows: $G_{\text{critical}} = (n-1)t_{\text{critical}} / \sqrt{n}$. In this equation, n represents the total number of observations in the dataset, and t_{critical} is the value obtained from the **t-distribution**.

The value of t_{critical} is determined based on the degrees of freedom, defined as $n - 2$, and the chosen **significance level** (often denoted as alpha, α). For a single-tailed test, the significance level used in the t-lookup is α/n , whereas for a two-tailed test, the level is adjusted to $\alpha/(2n)$. This adjustment is a form of the **Bonferroni correction**, which accounts for the fact that we are testing the most extreme value among n observations, thereby maintaining the overall **Type I error** rate at the desired level.

In **Excel**, calculating this **critical value** requires the use of the `T.INV.2T` or `T.INV` functions. By inputting the adjusted alpha and the degrees of freedom, **Excel** can provide the precise t-value needed to complete the **Grubbs' Test** formula. If the calculated G statistic exceeds this **critical value**, we conclude with a specific level of confidence--usually 95% or 99%--that the data point in question is a legitimate **outlier** that warrants further investigation or removal.

Step 1: Visualizing Data and Verifying Normality

Before proceeding with numerical calculations, it is an essential best practice in **statistics** to visualize the data. The **Grubbs' Test** assumes that the underlying population from which the sample is drawn follows a **normal distribution**. If the data is heavily skewed or follows a different distribution type, the results of the test may be invalid. A **histogram** is the most effective tool for this preliminary check, allowing the analyst to see if the data roughly follows a symmetrical, bell-shaped curve.

To create a **histogram** in **Excel**, select your data range and navigate to the "Insert" tab, then choose the Statistical Chart icon to insert the histogram. This visual representation provides immediate insight into the spread of the data and the presence of any values that appear far

removed from the central cluster. In the example below, we are investigating whether the value 60 stands out significantly from the rest of the observations in the provided list.

	A	B	C	D
1	Raw Data			
2	5			
3	10			
4	12			
5	14			
6	15			
7	15			
8	14			
9	13			
10	19			
11	17			
12	16			
13	20			
14	22			
15	8			
16	21			
17	28			
18	11			
19	9			
20	29			
21	60			
22				

The following screenshots demonstrate the process of generating the **histogram** within the **Excel** environment. Notice how the chart groups the data into "bins," which helps in identifying the frequency and distribution of values. If the **histogram** shows a relatively balanced distribution with a peak in the center, the assumption of normality is likely satisfied, and you may proceed with the **Grubbs' Test** with confidence.

	A	B	C	D	E	F	G	H	I
1	Raw Data		bins						
2	5		5						
3	10		10						
4	12		15						
5	14		20						
6	15		25						
7	15		30						
8	14								
9	13								
10	19								
11	17								
12	16								
13	20								
14	22								
15	8								
16	21								
17	28								
18	11								
19	9								
20	29								
21	60								

Histogram ? X

Input

Input Range:

Bin Range:

Labels

Output options

Output Range:

New Worksheet Ply:

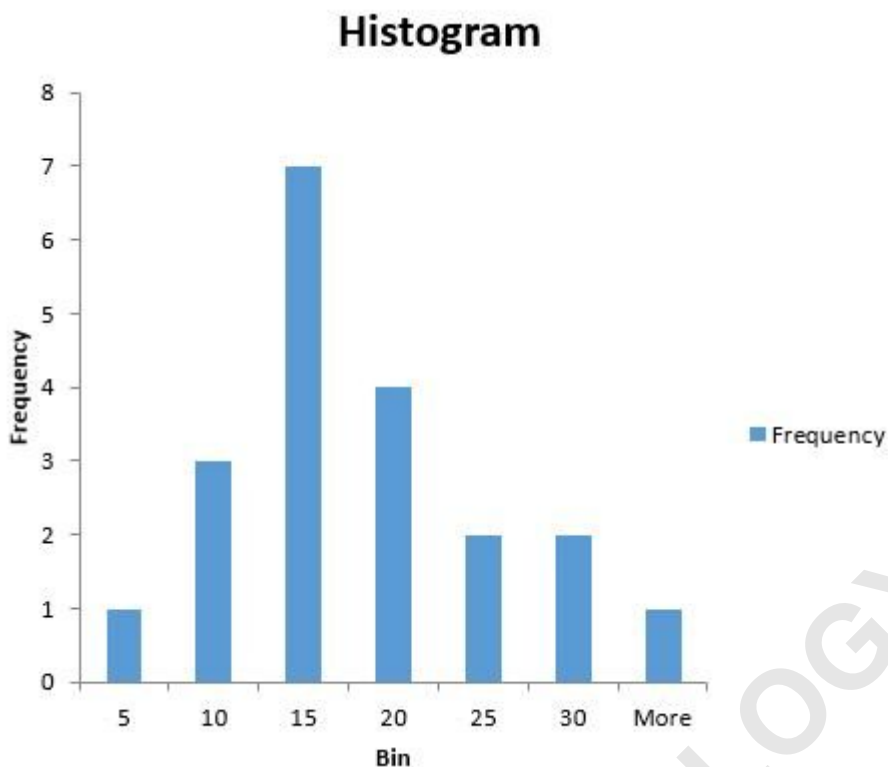
New Workbook

Pareto (sorted histogram)

Cumulative Percentage

Chart Output

OK Cancel Help



Step 2: Executing Grubbs' Test Calculations in Excel

After verifying the **normal distribution** assumption, the next step is to perform the actual calculations for the **Grubbs' Test**. This involves several sub-steps in **Excel**: calculating the **sample mean**, determining the **standard deviation**, identifying the maximum or minimum value, and finally computing the **G** statistic and the **critical value**. Utilizing **Excel** formulas like ``AVERAGE()``, ``STDEV.S()``, and ``MAX()`` makes this process efficient and reduces the likelihood of manual calculation errors.

In our specific example, we are testing the value 60. By organizing the calculations in a clear, labeled format within the spreadsheet, you can easily see the relationship between the various components. The screenshot below illustrates the layout and the specific formulas required to derive both the test statistic and the **critical value**. This structured approach is highly recommended for maintaining a transparent and auditable **data analysis** workflow.

	A	B	C	D	E
1	Raw Data		Max	60	=MAX(A2:A21)
2	5		Mean	17.9	=AVERAGE(A2:A21)
3	10		Standard Deviation	11.684	=STDEV.S(A2:A21)
4	12		G	3.603219	=(D1-D2)/D3
5	14				
6	15		Alpha	0.05	
7	15		Sample Size	20	=COUNT(A2:A21)
8	14		Significance Value	0.0025	=D6/D7
9	13		Degrees of Freedom	18	=D7-2
10	19		t-Critical Value	3.196574	=T.INV(1-D8, D9)
11	17		G-Critical Value	2.556581	=(D7-1)*D10/SQRT(D7*(D9+D10^2))
12	16				
13	20				
14	22				
15	8				
16	21				
17	28				
18	11				
19	9				
20	29				
21	60				
22					

As shown in the analysis, the test statistic **G** located in cell D4 is calculated as **3.603219**. Simultaneously, the **critical value Gcritical** calculated in cell D11 is **2.556581**. Because the observed **G** value is substantially higher than the **critical value**, we have strong statistical evidence to suggest that the value 60 is an **outlier**. This objective result removes the guesswork from **outlier** detection, providing a mathematical justification for any subsequent data adjustments.

Strategic Management of Identified Outliers

When the **Grubbs' Test** successfully identifies an **outlier** in your dataset, the next challenge is deciding how to handle that specific data point. Identifying an **outlier** is merely a diagnostic step; the remedy depends heavily on the context of the research and the likely cause of the anomaly. The first and most critical step is to perform a thorough investigation to ensure the value is not the result of a simple **data entry** error or a clerical mistake. Often, extreme values are merely typos that can be easily corrected by referencing original data sources.

If the **outlier** is confirmed to be an error but the correct value is unrecoverable, you may choose to assign a new value to it--a process sometimes referred to as **imputation** or **Winsorization**. This

might involve replacing the **outlier** with the **sample mean** or the next nearest non-outlying value. However, this approach must be used with caution, as it can artificially reduce the **variance** of the dataset and potentially bias the results of subsequent **statistical analysis**.

Alternatively, if the value is a legitimate measurement but represents a rare or extreme event that does not align with the goals of your study, you may choose to remove the **outlier** entirely. This is common when the anomaly is suspected to be the result of equipment malfunction or environmental interference. Regardless of the chosen path, it is ethically imperative to document the detection and subsequent treatment of outliers in your final report. Transparency regarding **outlier** management ensures that other researchers can accurately interpret your findings and reproduce your results.

Summary of Best Practices for Outlier Detection

Conducting a **Grubbs' Test** in **Excel** is a powerful way to enhance the **reliability** of your data. To ensure the best results, always follow a systematic workflow: start with **data visualization**, verify **normality**, use precise formulas for the **G** statistic and **critical value**, and conclude with a thoughtful interpretation of the results. By mastering these steps, you transform **Excel** from a simple spreadsheet tool into a robust environment for professional-grade **statistical analysis**.

The following list summarizes the key actions to take when managing potential anomalies:

Verify Data Accuracy: Always check if the **outlier** is a result of a typo or **data entry** error by cross-referencing with primary sources before taking statistical action.

Evaluate the Cause: Determine if the **outlier** was caused by an identifiable error, such as a faulty sensor or a miscalibrated instrument, which justifies removal.

Assess Impact: Consider how the presence or absence of the **outlier** changes the **sample mean** and **standard deviation**, and whether it alters the final conclusions of the study.

Document Methodology: Clearly state in your **data analysis** report that a **Grubbs' Test** was performed, including the **significance level** used and the justification for removing or keeping any identified points.

In conclusion, the **Grubbs' Test** provides a rigorous mathematical framework for **outlier** detection that far exceeds the accuracy of simple "eye-balling" the data. When implemented correctly in **Excel**, it serves as a cornerstone for maintaining **data quality** and ensuring that your statistical models are built on a foundation of accurate and representative observations. Whether you are working in finance, engineering, or social sciences, these techniques are essential for any professional **data analyst**.