

# Calculate Standardized Residuals in Excel

Authored by  
**Mohammed loot**

November 6, 2025

## RECOMMENDED CITATION

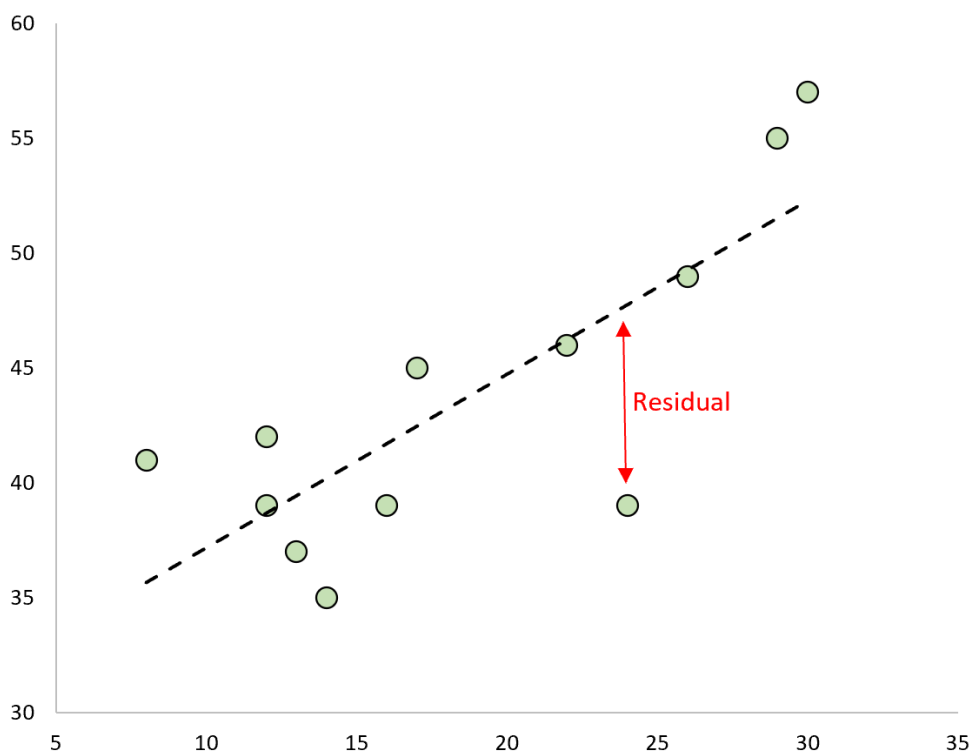
Mohammed loot (2025). *Calculate Standardized Residuals in Excel*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=11463>

In statistical analysis, understanding the difference between observed data points and the values predicted by a model is fundamental. This difference is known as a **residual**. Specifically, within a **regression model**, the residual quantifies the error for a given observation.

The calculation of a residual is straightforward:

**Residual = Observed value - Predicted value**

When plotting observed data points alongside the fitted regression line, the residual for each point represents the precise vertical distance between that observation and the line of best fit.



## The Importance of Standardized Residuals in Model Diagnostics

While simple residuals are useful, they do not account for the varying influence or **leverage** that different observations might have on the regression line. This is where the **standardized residual** becomes a crucial diagnostic tool, especially for identifying unusual data points or **outliers** in the model.

A standardized residual is calculated by dividing the raw residual by its standard deviation, adjusting for the leverage of that observation. The formula is defined as follows:

$$r_i = e_i / s(e_i) = e_i / \text{RSE} \sqrt{1 - h_{ii}}$$

The components of this formula are essential for accurate interpretation:

**$e_i$** : The  $i$ th raw residual (Observed  $Y$  minus Predicted  $Y$ ).

**RSE**: The [Residual Standard Error](#) of the entire regression model.

**$h_{ii}$** : The leverage of the  $i$ th observation.

Statisticians typically use standardized residuals to flag potential outliers. A common rule of thumb suggests that any standardized residual with an absolute value greater than 3 merits investigation as a serious outlier. This tutorial will walk you through the practical steps of performing this complex calculation using Microsoft Excel.

## Step 1: Entering and Organizing the Dataset

To begin the process in Excel, we must first input our dataset. For this example, we will use a small collection of paired ( $X$ ,  $Y$ ) values. Ensure your independent variables ( $X$ ) and dependent variables ( $Y$ ) are clearly labeled and separated into adjacent columns.

Here is the sample data organized in the spreadsheet:

	A	B	C	D	E	F	G
1		<b>X</b>	<b>Y</b>				
2		8	41				
3		12	42				
4		12	39				
5		13	37				
6		14	35				
7		16	39				
8		17	45				
9		22	46				
10		24	39				
11		26	49				
12		29	55				
13		30	57				
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							

This initial organization is crucial for the successful execution of the regression analysis in the next step.

## Step 2: Running Regression and Calculating Raw Residuals

The raw residuals are automatically generated when running a regression analysis using Excel's built-in tools. To access these tools, navigate to the **Data** tab located on the top ribbon. Within the **Analysis** group, click on **Data Analysis**.

If the Data Analysis option is not visible, you must install the [Analysis ToolPak](#) Add-in. This is a free and straightforward process that enables advanced statistical functions in Excel.

When the Data Analysis window appears, select the **Regression** option and click **OK**. You will then need to configure the following inputs in the subsequent dialogue box before clicking **OK** again:

Input Y Range: Select the column containing your dependent variable (Y).

Input X Range: Select the column(s) containing your independent variable(s) (X).

Crucially, check the box labeled **Residuals** to ensure the raw residuals are included in the output report.

	A	B	C	D	E	F	G	H	I	J
1		X	Y							
2		8	41							
3		12	42							
4		12	39							
5		13	37							
6		14	35							
7		16	39							
8		17	45							
9		22	46							
10		24	39							
11		26	49							
12		29	55							
13		30	57							
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										

Regression

Input

Input Y Range: \$C\$1:\$C\$13

Input X Range: \$B\$1:\$B\$13

Labels  Constant is Zero

Confidence Level: 95 %

Output options

Output Range: \$B\$21

New Worksheet Ply:

New Workbook

Residuals

Residuals  Residual Plots

Standardized Residuals  Line Fit Plots

Normal Probability

Normal Probability Plots

OK Cancel Help

Upon execution, the regression output will be generated, including a dedicated section detailing the raw residuals for every data point.

40				
41				
42		RESIDUAL OUTPUT		
43				
44		<i>Observation</i>	<i>Predicted Y</i>	<i>Residuals</i>
45		1	35.6732	5.326796
46		2	38.69435	3.305645
47		3	38.69435	0.305645
48		4	39.44964	-2.44964
49		5	40.20493	-5.20493
50		6	41.71551	-2.71551
51		7	42.47079	2.529206
52		8	46.24723	-0.24723
53		9	47.75781	-8.75781
54		10	49.26838	-0.26838
55		11	51.53425	3.465752
56		12	52.28954	4.710464
57				
58				
59				
60				
61				
62				

For ease of calculation in the subsequent steps, copy and paste these raw residual values into a new, clearly labeled column adjacent to your original data.

	A	B	C	D	E
1		<b>X</b>	<b>Y</b>	<b>Residuals</b>	
2		8	41	5.326796	
3		12	42	3.305645	
4		12	39	0.305645	
5		13	37	-2.44964	
6		14	35	-5.20493	
7		16	39	-2.71551	
8		17	45	2.529206	
9		22	46	-0.24723	
10		24	39	-8.75781	
11		26	49	-0.26838	
12		29	55	3.465752	
13		30	57	4.710464	
14					
15					
16					
17					
18					
19					
20					

### Step 3: Calculating the Leverage (hii) for Each Observation

To calculate the standardized residual, we must first determine the leverage (**hii**) for each observation. Leverage measures how far an observation's predictor values are from the mean of the predictor values, indicating its potential influence on the regression line. For simple linear regression (one X variable), the leverage calculation is as follows:  $h_{ii} = 1/n + (x_i - \bar{x})^2 / \sum(x_j - \bar{x})^2$ .

To implement this in Excel, we first need to calculate three auxiliary values based on the independent variable (X): the count (n), the mean ( $\bar{x}$ ), and the sum of squared deviations ( $\sum(x_j - \bar{x})^2$ ).

	A	B	C	D	E	F
1		<b>X</b>	<b>Y</b>	<b>Residuals</b>	<b>Leverage</b>	
2		8	41	5.326796	0.271607	
3		12	42	3.305645	0.156184	
4		12	39	0.305645	0.156184	
5		13	37	-2.44964	0.135733	
6		14	35	-5.20493	0.118644	
7		16	39	-2.71551	0.094551	
8		17	45	2.529206	0.087547	
9		22	46	-0.24723	0.102956	
10		24	39	-8.75781	0.132652	
11		26	49	-0.26838	0.175795	
12		29	55	3.465752	0.265723	
13		30	57	4.710464	0.302423	
14	<b>n</b>	12				
15	<b>Mean</b>	18.583				
16	<b>SS</b>	594.917				
17						
18						
19						
20						

The following formulas were used in the cells shown above to derive the necessary components and the final leverage values (assuming X data is in B2:B13):

**B14 (n):** =COUNT(B2:B13) (Counts the number of observations)

**B15 (Mean X):** =AVERAGE(B2:B13) (Calculates the mean of X)

**B16 (Sum of Squared Deviations):** =DEVSQ(B2:B13) (Calculates  $\sum(x_j - \bar{x})^2$ )

**E2 (Leverage hii):** =1/\$B\$14+(B2-\$B\$15)^2/\$B\$16 (This formula is then dragged down for all observations)

#### Step 4: Deriving the Standardized Residuals

With the raw residuals ( $e_i$ ) and the leverage values ( $h_{ii}$ ) calculated, we must now incorporate the **Residual Standard Error** (RSE) to find the final standardized residual ( $r_i$ ).

The formula for the standardized residual is:

$$r_i = e_i / \text{RSE} \sqrt{1-h_{ii}}$$

The RSE represents the average distance that the observed values fall from the regression line and is extracted directly from the summary output generated in Step 2. In our example, the RSE value is listed as the "Standard Error" and equals **4.44**.

13		30	57	4.710464	0.302423
14	<b>n</b>	12			
15	<b>Mean</b>	18.583			
16	<b>SS</b>	594.917			
17					
18					
19					
20					
21		SUMMARY OUTPUT			
22					
23		<i>Regression Statistics</i>			
24		Multiple R	0.795222		
25		R Square	0.632378		
26		Adjusted R Square	0.595615		
27		Standard Error	4.441741		
28		Observations	12		
29					
30		ANOVA			
31			<i>df</i>	<i>SS</i>	<i>MS</i>
32		Regression	1	339.376	339.376
33		Residual	10	197.2907	19.72907
34		Total	11	536.6667	
35					
36			<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
37		Intercept	29.6309	3.618911	8.187794
38		X	0.755288	0.182106	4.147509
39					

Using this constant RSE value (4.44) and the calculated raw residuals (Column C) and leverage values (Column E), we can construct the final Excel formula for the standardized residual in Column F:

	A	B	C	D	E	F
1		X	Y	Residuals	Leverage	Standardized Residuals
2		8	41	5.326796	0.271607	=D2/(4.44*SQRT(1-E2))
3		12	42	3.305645	0.156184	0.810
4		12	39	0.305645	0.156184	0.075
5		13	37	-2.44964	0.135733	-0.593
6		14	35	-5.20493	0.118644	-1.249
7		16	39	-2.71551	0.094551	-0.643
8		17	45	2.529206	0.087547	0.596
9		22	46	-0.24723	0.102956	-0.059
10		24	39	-8.75781	0.132652	-2.118
11		26	49	-0.26838	0.175795	-0.067
12		29	55	3.465752	0.265723	0.911
13		30	57	4.710464	0.302423	1.270
14	n	12				
15	Mean	18.583				
16	SS	594.917				
17						
18						
19						
20						

## Step 5: Interpreting Results and Identifying Outliers

The final column provides the standardized residual for every observation. These values allow for an easy assessment of how unusual each data point is, independent of its leverage.

Based on the calculated results, we observe that none of the standardized residuals have an absolute value exceeding 3. Therefore, under the widely accepted statistical threshold, there is no evidence to suggest that any observation in this dataset is a significant **outlier**.

It is important to note that while an absolute value of 3 is a common benchmark for outlier detection, some researchers prefer a more conservative threshold, such as an absolute value of 2. The choice between these thresholds often depends on the specific domain, the size of the dataset, and the required sensitivity of the statistical analysis.

## Additional Resources

[What Are Standardized Residuals?](#)