

# How to Calculate Standardized Residuals in Excel?

Authored by  
**stats writer**

December 15, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Calculate Standardized Residuals in Excel?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=107477>

Standardized residuals are a fundamental tool in statistical analysis, particularly within the context of regression models. They provide a critical measure for assessing the reliability of your model and identifying observations that deviate significantly from the predicted trend. Effectively, they standardize the traditional residual, allowing for easier comparison across different datasets and for robust detection of influential data points or outliers.

This detailed guide will walk you through the precise, multi-step process of calculating these powerful metrics entirely within Microsoft Excel. While dedicated statistical software often handles this automatically, understanding the manual calculation in Excel provides deeper insight into the underlying statistical principles. We will break down the required steps, focusing on using Excel's Data Analysis ToolPak and applying complex statistical formulas for high precision.

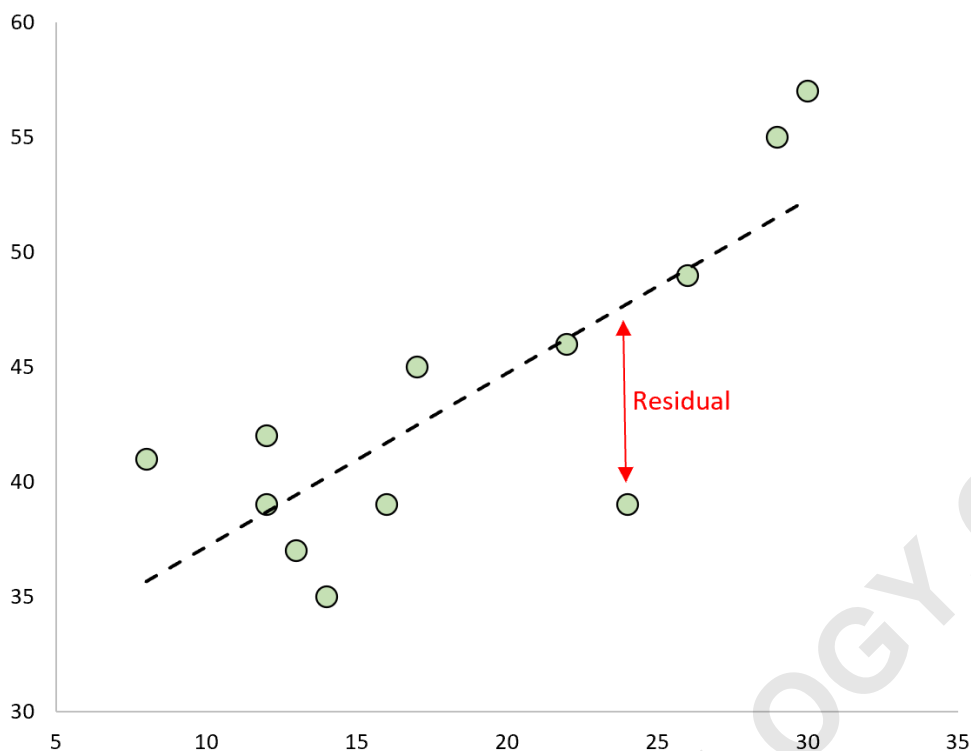
## The Foundational Concept of a Residual

A **residual** represents the fundamental difference between an observed data value (Y) and the value predicted by the regression model (?). This raw difference quantifies the error inherent in the model's prediction for a specific observation. A small residual indicates a close fit, while a large residual suggests the model poorly predicts that particular data point.

The standard calculation for the error term is:

**Residual = Observed value - Predicted value**

When visualizing a scatter plot with an overlaid fitted regression line, the residual for any given point is the exact vertical distance between that observation and the regression line. This visual representation clearly shows how far each observation deviates from the estimated relationship described by the model.



## Understanding Standardized Residuals and Outlier Detection

While standard residuals are informative, their magnitude depends entirely on the scale of the original data. To objectively compare these errors and reliably identify extreme observations--especially when the variance of the residuals might not be constant (heteroscedasticity)--we use the **standardized residual**. This metric scales the residual by an estimate of its standard deviation.

The standardized residual is superior for model diagnostics because it accounts for both the overall variation in the errors and the influence that high-leverage points might have on the calculation. It effectively transforms the error into a unitless quantity that behaves much like a Z-score, making the identification of statistically significant outliers much more reliable and comparable across various studies.

The formal calculation for the standardized residual ( $r_i$ ) is defined as:

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

Where the statistical components are critical:

**$e_i$** : The  $i$ th raw residual.

**RSE**: The Residual Standard Error of the model. This is the estimate of the standard deviation of

the true error term ( $\sigma$ ).

**hii:** The leverage of the  $i$ th observation.

For practical purposes, a widely used statistical convention suggests that any standardized residual with an absolute value greater than 3 should be flagged as a potential, significant outlier, demanding careful review.

## Step 1: Data Entry and Preparation in Excel

The process begins with accurate data setup. For this manual calculation in Excel, we assume a simple linear regression scenario where we have one independent variable (X) and one dependent variable (Y). Ensure your data is clean, complete, and organized into two contiguous, labeled columns.

We will utilize the following sample dataset to demonstrate the entire calculation pipeline, laying the groundwork for the subsequent statistical computations.

	A	B	C	D	E	F	G
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							

Accurate input ranges are critical for the next step, where Excel's regression tool will calculate the raw residuals and the overall model statistics.

## Step 2: Generating Raw Residuals and RSE via Regression

To calculate the raw residuals ( $e_i$ ) and obtain the essential Residual Standard Error (RSE), we must run a standard linear regression analysis using the Excel Data Analysis ToolPak. Ensure this add-in is enabled (File > Options > Add-ins).

Access the tool by navigating to the **Data** tab, clicking **Data Analysis**, and selecting **Regression**. In the dialog box, specify the Input Y Range (Dependent Variable) and the Input X Range (Independent Variable). Most importantly, you must check the box for **Residuals** under the Residuals section.

	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:  ↑

Input X Range:  ↑

Labels  Constant is Zero

Confidence Level:  %

Output options

Output Range:  ↑

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 output will contain a summary of the regression statistics and a dedicated table showing the output residuals, which are the raw differences between the observed and predicted Y values.

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 efficient subsequent calculations, copy the raw residuals from this output and paste them into a new column right next to your original dataset. This columnar setup streamlines the application of the standardized residual formula.

	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 Statistic ( $h_{ii}$ )

The leverage statistic ( $h_{ii}$ ) is necessary because the variance of the raw residual is not constant; it depends on the distance of the predictor value ( $x_i$ ) from the mean of the predictors ( $\bar{x}$ ). High-leverage points are those whose X values are extreme, potentially pulling the regression line towards them.

For a simple linear regression, the leverage formula is based on the sample size ( $n$ ) and the squared deviations of the X values:

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2}$$

To calculate this in Excel, we first summarize the X variable (the values in column B). We need the count, the average, and the sum of squared deviations.

	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 required summary formulas are:

**Count (\$n\$):** =COUNT(B2:B13) (In cell B14)

**Mean ( $\bar{x}$ ):** =AVERAGE(B2:B13) (In cell B15)

**Sum of Squared Deviations:** =DEVSQ(B2:B13) (In cell B16)

Finally, the leverage for the first observation (in cell E2) is calculated using absolute references for the summary statistics: =1/\$B\$14+(B2-\$B\$15)^2/\$B\$16. Drag this formula down column E to complete the leverage calculation for all observations.

#### Step 4: Compiling the Standardized Residuals

The final step involves integrating the raw residual ( $e_i$ ), the RSE, and the leverage ( $h_{ii}$ ) into the standardization formula. We must first retrieve the RSE (which is constant for the entire model) from the output generated in Step 2.

Locate the RSE, typically labeled as "Standard Error" within the Regression Statistics. In our example, the RSE value is **4.44**. It is advisable to place this value in a fixed cell (e.g., \$G\$1) for easy referencing.

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>	<i>F</i>	
32		Regression	1	339.376	339.376	17.2018	
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>	<i>P-value</i>
37		Intercept	29.6309	3.618911	8.187794	9.6038	
38		X	0.755288	0.182106	4.147509	0.00198	
39							

The standardization formula requires dividing the raw residual by the adjusted standard error of that residual:

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

If the raw residual ( $e_i$ ) is in column D and the leverage ( $h_{ii}$ ) is in column E, the complete Excel formula for the standardized residual in cell F2 is:

$$=D2 / (\$G\$1 * \text{SQRT}(1-E2))$$

Copying this formula down column F yields the complete set of standardized residuals for the dataset:

	A	B	C	D	E	F
1		<b>X</b>	<b>Y</b>	<b>Residuals</b>	<b>Leverage</b>	<b>Standardized Residuals</b>
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	<b>n</b>		12			
15	<b>Mean</b>		18.583			
16	<b>SS</b>		594.917			
17						
18						
19						
20						

## Interpreting the Diagnostic Results

The final column of standardized residuals provides the crucial diagnostic information. Since these values are standardized, we can directly compare them to common thresholds for identifying statistical outliers.

As a standard benchmark, observations where the absolute value of the standardized residual exceeds 3 are typically considered significant outliers. Examining the results from our calculation, all standardized residuals fall between -3 and 3. This indicates that, statistically, none of the individual observations are exerting undue influence or represent extreme errors relative to the variability of the model.

It is worth noting that some statisticians may employ a more conservative threshold, such as an absolute value of 2, particularly in domains where high sensitivity to anomalies is required. The choice of threshold (2 or 3) should be determined by the requirements of the specific research context and must be applied consistently.

### What Are Standardized Residuals?