

How to Calculate Residual Sum of Squares in Excel

Authored by
stats writer

December 13, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Calculate Residual Sum of Squares in Excel*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=107374>

The Residual Sum of Squares (RSS), often referred to as the sum of squared errors (SSE), is a fundamental metric used in regression analysis to evaluate the goodness-of-fit of a statistical model. Essentially, RSS quantifies the disparity between the observed data points and the values predicted by the chosen regression line or curve. This metric is derived by calculating the difference between each actual data point and its corresponding predicted value on the regression line, squaring these differences to eliminate negative values, and then aggregating all the squared discrepancies. A lower RSS value indicates a model that fits the underlying data structure more accurately. Calculating this crucial value is readily achievable in Microsoft Excel, typically utilizing advanced array functions like SUMPRODUCT in combination with the POW function, or, more conveniently, through the powerful array statistical function, LINEST.

Understanding the RSS is paramount for anyone engaging in data modeling, as it serves as an absolute measure of the error variance inherent in the data set that the independent variables fail to explain. While simple calculation methods involving manual prediction and subtraction are possible, leveraging Excel's built-in statistical functionalities streamlines the process significantly, allowing users to rapidly assess the efficacy of various regression models, whether they involve a single predictor (simple linear regression) or multiple predictors (multiple linear regression). The techniques detailed below focus on utilizing the LINEST function, which provides a comprehensive statistical summary of the regression model, including the RSS, without requiring intermediate calculations of predicted Y values.

Defining the Residual and Sum of Squares

Before diving into the calculation of the Residual Sum of Squares, it is essential to establish a clear definition of the fundamental component: the residual. In the context of any regression model, a residual represents the vertical distance between an observed data point (the actual value of the dependent variable) and the corresponding point on the fitted regression line (the predicted value). This difference captures the error or noise in the data that the model fails to account for. If a data point lies exactly on the regression line, its residual is zero, indicating a perfect prediction at that specific instance.

Mathematically, the residual for the i -th observation, often denoted as e_i , is straightforwardly calculated as the difference between the observed value (Y_i) and the predicted value (\hat{Y}_i):

Residual (e_i) = Observed value (Y_i) - Predicted value (\hat{Y}_i)

The Residual Sum of Squares (RSS) takes these individual residuals and squares them, a crucial step that ensures both positive and negative errors contribute positively to the total error measure, and simultaneously magnifies the impact of larger errors (outliers). Subsequently, all these squared residuals are summed up to yield the final RSS value, which acts as a consolidated measure of the

overall prediction error of the fitted line relative to the observed data points.

The formal equation for the Residual Sum of Squares is defined as:

$$\text{Residual Sum of Squares (RSS)} = \sum(e_i)^2$$

Where the constituent parts of this formula are defined as follows:

Σ : This is the Greek capital letter sigma, representing the mathematical operation of summation, signifying that we must sum all subsequent terms.

e_i : This term denotes the i th residual, which is the error associated with the i th data point.

This resulting value is fundamental for statistical interpretation: the lower the magnitude of the RSS, the closer the predicted values are to the actual observed values, signifying a better-fitting and more accurate regression model. This tutorial focuses on efficient methods within Microsoft Excel to quickly determine this critical statistical measure for both simple linear regression and multiple linear regression models.

Prerequisites: Understanding Excel's LINEST Function

While one could calculate the RSS manually by first determining the regression coefficients, calculating predicted Y values, finding the residuals, squaring them, and summing them up, Excel provides a much more streamlined approach using the LINEST function. LINEST is an array function designed specifically to calculate the statistics for a line using the least squares method, which is the standard optimization technique for fitting a regression line. This function returns an array of values, including the slope(s), intercept, and various statistical metrics--one of which is the RSS.

To utilize the LINEST function effectively, especially when requesting a full set of regression statistics, it must be entered as an array formula across a range of cells, typically five rows high and equal to the number of predictor variables plus one column wide. The syntax is crucial for successful deployment:

LINEST(known_ys, , ,)

Each argument plays a specific role in defining the data and the required output:

known_ys: This is a required argument specifying the range of cells containing the dependent variable (Y) values. These are the observed values you are trying to predict.

known_xs: This is an optional argument specifying the range of cells containing the independent variable(s) (X) values. If omitted, Excel assumes the X values are the array {1, 2, 3, ...} corresponding to the Y values.

const: This is an optional logical value. If set to TRUE (or omitted), the intercept (constant b) is calculated normally. If set to FALSE, the intercept b is forced to be equal to zero, meaning the regression line must pass through the origin. For standard regression, we typically leave this blank or set it to TRUE.

stats: This is an optional logical value. If set to TRUE, the function returns the comprehensive regression statistics array, which includes the Residual Sum of Squares. If FALSE (or omitted), only the coefficients and intercept are returned. We must specify this as TRUE to obtain the RSS.

Example 1: Calculating RSS for Simple Linear Regression

Let us consider a practical application using a dataset suitable for simple linear regression, where we have a single predictor variable (X) and one response variable (Y). Suppose we have the following sample dataset entered into columns A and B of an Excel worksheet, representing various pairs of observations:

	A	B	C	D	E	F
1	x	y				
2	6	16				
3	7	17				
4	7	17				
5	8	19				
6	12	22				
7	14	24				
8	15	25				
9	16	27				
10	16	33				
11	19	34				
12	21	35				
13	22	35				
14	23	34				
15	24	36				
16	24	39				
17						
18						
19						
20						
21						
22						
23						
24						
25						

To calculate the Residual Sum of Squares for a simple linear model where X is the predictor and Y is the response, we must instruct the LINEST function to return the extended statistics array. Since

there is one predictor, the output array will be two columns wide (for the slope and intercept) and five rows deep. We must select a range of 10 cells (e.g., C2:D6) before entering the formula.

The specific formula used in this context, assuming Y values are in B2:B11 and X values are in A2:A11, is:

```
{=LINEST(B2:B11, A2:A11, TRUE, TRUE)}
```

Note that the curly braces indicate that this formula must be entered as an array formula (by pressing Ctrl+Shift+Enter in older versions of Excel, or simply Enter in modern versions supporting dynamic arrays). The final argument, TRUE, ensures that the comprehensive statistical output is generated, which is necessary to locate the RSS value.

Interpreting the LINEST Function Output

Once the LINEST function successfully executes as an array formula, it populates the selected range with a 5x2 matrix of statistical information for simple linear regression. This output array is standardized, and the location of the RSS is fixed. The output structure typically looks like this (from top left to bottom right): coefficients (slopes), standard errors, R-squared and standard error for the Y-estimate, F-statistic and degrees of freedom, and finally, the Regression Sum of Squares (SSR) and the Residual Sum of Squares (RSS).

The following visual representation confirms the application of the LINEST array formula and displays the resulting statistical matrix:

	A	B	C	D	E	F	G	H	I
1	x	y							
2	6	16		1.202164	8.779581		=LINEST(B2:B16, A2:A16, , TRUE)		
3	7	17		0.081237	1.366128				
4	7	17		0.943963	1.975904				
5	8	19		218.9896	13				
6	12	22		854.9788	50.75456				
7	14	24							
8	15	25							
9	16	27							
10	16	33							
11	19	34							
12	21	35							
13	22	35							
14	23	34							
15	24	36							
16	24	39							
17									
18									
19									
20									
21									
22									
23									

As clearly shown in the resulting array, the Residual Sum of Squares (RSS) for this specific regression model is consistently located in the bottom-right cell of the 5x2 output matrix. In this particular numerical example derived from the dataset, the calculated residual sum of squares is determined to be exactly **50.75**. This value quantifies the total unexplained variance in Y after accounting for the linear relationship with the predictor X.

Example 2: Calculating RSS for Multiple Linear Regression

The process for calculating the Residual Sum of Squares for a multiple linear regression model is fundamentally identical to the simple linear case, relying entirely on the robust LINEST function. The primary difference lies in the configuration of the input data, specifically the **known_xs** argument, which now encompasses multiple columns corresponding to the several independent predictor variables. Suppose we introduce a second predictor variable, X2, leading to the following dataset structure:

	A	B	C	D	E	F	G
1	x1	x2	y				
2	6	8	16				
3	7	9	17				
4	7	11	17				
5	8	11	19				
6	12	11	22				
7	14	12	24				
8	15	12	25				
9	16	12	27				
10	16	13	33				
11	19	13	34				
12	21	14	35				
13	22	15	35				
14	23	15	34				
15	24	17	36				
16	24	19	39				
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							

In this scenario, we are modeling Y as a function of both X1 and X2. Because we have two predictor variables, the output array generated by LINEST will be wider: three columns wide (one for the coefficient of X1, one for X2, and one for the intercept) and still five rows deep to accommodate all standard statistics. Consequently, we must select a 5x3 range of cells (e.g., D2:F6) prior to entering the formula to ensure the entire statistical output is captured.

The key modification in the LINEST formula is the inclusion of both predictor columns in the range definition for **known_xs**. Assuming Y is in A2:A11 and predictors X1 and X2 are in B2:C11, the formula entered as an array would be:

```
{=LINEST(A2:A11, B2:C11, TRUE, TRUE)}
```

The formula specifies the continuous range B2:C11, effectively treating both columns as the collective set of independent variables for the multiple linear regression. The final arguments, TRUE and TRUE, remain necessary to calculate the intercept and display the full statistical matrix,

respectively.

	A	B	C	D	E	F	G	H	I	J	K
1	x1	x2	y								
2	6	8	16		0.22402	1.11098	7.334584		=LINEST(C2:C16, A2:B16, , TRUE)		
3	7	9	17		0.475237	0.210801	3.373707				
4	7	11	17		0.944982	2.037806	#N/A				
5	8	11	19		103.0548	12	#N/A				
6	12	11	22		855.9015	49.83183	#N/A				
7	14	12	24								
8	15	12	25								
9	16	12	27								
10	16	13	33								
11	19	13	34								
12	21	14	35								
13	22	15	35								
14	23	15	34								
15	24	17	36								
16	24	19	39								
17											
18											
19											
20											
21											
22											
23											
24											

After executing the formula, the Residual Sum of Squares remains fixed in its positional location: the bottom-right corner of the resulting statistical matrix, regardless of the number of predictors. In this instance involving multiple linear regression, the calculated residual sum of squares turns out to be **49.83**. The marginal reduction in the RSS (from 50.75 in Example 1 to 49.83 in Example 2) suggests that the inclusion of the second predictor variable, X2, slightly improved the overall fit of the regression model to the observed data.

Advanced Interpretation: Why a Lower RSS Matters

The magnitude of the Residual Sum of Squares is a direct indicator of the spread or variability of the observed data points around the fitted regression line. A low RSS suggests that the model's predictions are very close to the actual outcomes, implying the independent variables successfully capture most of the variance in the dependent variable. Conversely, a high RSS means there is a significant amount of unexplained error, often pointing towards a poorly specified model, missing relevant predictor variables, or inherent noise in the data that cannot be linearly modeled.

In practical terms, the RSS is crucial because it forms the basis for calculating several other important statistical measures, including the standard error of the estimate (which is derived from the mean square error, RSS divided by its degrees of freedom). Furthermore, when comparing

different regression models built on the same dataset (as seen between Example 1 and Example 2), researchers often look for the model that minimizes the RSS, assuming all other statistical checks (like handling overfitting and ensuring assumptions are met) are satisfied. The ability to quickly and accurately extract this value using Microsoft Excel's [LINEST function](#) significantly expedites the iterative process of model selection and refinement in data analysis.

ARABPSYCHOLOGY.COM