

Calculate Residual Sum of Squares in Excel

Authored by
Mohammed loot

November 6, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Calculate Residual Sum of Squares in Excel*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=11375>

The Foundation of Fit: Residuals and Statistical Modeling

In the world of statistical modeling, especially within the discipline of [regression analysis](#), a core concept that determines the quality of a predictive model is the **residual**. A residual represents the vertical distance--the error--between a specific observed data point and the value predicted by the established regression model. It is, fundamentally, a measure of how much the model missed the mark for that single observation.

Assessing these measurement errors is critically important for evaluating the effectiveness and predictive power of any statistical model. If a model were perfect--a hypothetical scenario rarely seen with real-world data--it would flawlessly capture the relationship between the independent and dependent variables, resulting in all residuals being exactly zero. Because natural data sets are inherently complex and prone to noise, residuals serve as a quantifiable statistical indicator of the model's deviation from the actual observed reality.

The calculation for any individual [residual](#) is straightforward, providing an immediate understanding of the model's local error at a given point:

Residual = Observed value - Predicted value

To move beyond single-point errors and assess the overall performance of the model across the entire dataset, statisticians rely on aggregate metrics. This necessity for an overall goodness-of-fit measure leads directly to the computation of the **residual sum of squares**, a cornerstone metric in model evaluation.

Understanding the Residual Sum of Squares (RSS) Metric

The **residual sum of squares** (RSS), sometimes alternatively known as the sum of squared errors (SSE), is an indispensable metric utilized to quantify the total discrepancy between the observed data points and the values predicted by the [regression analysis](#) line. It serves as a comprehensive indicator of the total unexplained variation remaining in the dependent variable after the model has been fitted.

The calculation process involves taking each individual residual, squaring it, and then summing these squared values across every observation within the dataset. The act of squaring the residuals is essential for two key reasons: first, it ensures that negative and positive errors do not cancel each other out, providing an accurate total magnitude of error; second, it applies a heavy penalty to larger errors, meaning that outliers or significant deviations have a disproportionately large impact on the final RSS value.

This mathematical procedure guarantees that the metric yields a positive, unambiguous value

representing the cumulative error. The formal mathematical representation of the **residual sum of squares** is given by the formula:

$$\text{Residual sum of squares} = \sum(e_i)^2$$

where the variables denote the following statistical components:

Σ : This is the uppercase Greek sigma symbol, representing the summation operator, which instructs the calculation to sum all subsequent terms.

e_i : This represents the i th residual, derived from the difference between the actual observed value (y_i) and the predicted value ($\hat{y}_i$).

The interpretation of the [residual sum of squares](#) is refreshingly intuitive: a lower RSS value signifies a smaller total squared error, indicating that the fitted model provides a closer and more accurate approximation of the observed data. Conversely, a high RSS suggests a poor fit, implying that the predictor variables used in the model fail to explain a substantial amount of the variance found in the response variable.

Essential Prerequisite: Leveraging the LINEST Function in Excel

While the calculation of RSS can theoretically be performed manually--by calculating individual residuals, squaring them, and then summing them--such an approach is tedious and error-prone, especially with large datasets. Fortunately, Microsoft Excel provides a powerful, built-in array function specifically designed to handle complex linear regression computations: the [LINEST function](#). This utility is paramount for analysts because it efficiently calculates the slope and intercept of the regression line and, crucially, returns a comprehensive statistical output matrix that includes the RSS.

The [LINEST function](#) operates as an array formula, meaning that it is capable of returning multiple values simultaneously across a specified range of cells. Depending on the version of Excel, it may require specific entry methods (such as pressing Ctrl+Shift+Enter in older editions) to activate the array output. The standardized syntax for utilizing this powerful statistical tool is structured as follows:

LINEST(known_ys, , ,)

Each argument controls the data input and the required level of statistical detail in the output:

known_ys: This is a required argument specifying the range of cells containing the dependent variable, or the Y-values that the model attempts to predict.

known_xs: This is an optional argument defining the range of cells containing the independent variable(s), or the predictor X-values. If this argument is omitted, Excel defaults to assuming the X-

values are the array {1, 2, 3, ...}.

const: This optional logical value (TRUE/FALSE) controls the intercept. Setting it to FALSE forces the constant b (the Y-intercept) to be zero. Typically, this is left blank or set to TRUE, allowing Excel to calculate the intercept freely.

stats: This crucial optional logical value (TRUE/FALSE) dictates the extent of the output. Setting this to **TRUE** instructs the function to return the full suite of regression statistics, including the R-squared value, standard errors, the F-statistic, and, most importantly for our purpose, the **residual sum of squares**.

When the *stats* argument is correctly set to TRUE, the [LINEST function](#) generates a five-row matrix with N columns (where N is the number of predictor variables plus one). Regardless of the complexity of the model, the **residual sum of squares** is always located in a fixed and predictable position within this output matrix: the last cell of the second column.

Practical Application 1: Calculating RSS for Simple Linear Regression

We will now walk through a detailed demonstration of utilizing the [LINEST function](#) to determine the **residual sum of squares** for a simple linear regression model, which, by definition, uses only a single predictor variable (X) to estimate a single response variable (Y).

Consider the following hypothetical dataset organized in an Excel worksheet, where we seek to establish a linear relationship to predict Y based on X:

	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						

Our primary goal is to execute the simple linear regression and retrieve the complete set of statistical outputs. To successfully accommodate the results, we must allocate an output range large enough for the 5-row by 2-column result matrix (since there is only one predictor variable).

The process requires following these precise steps:

Select an empty, contiguous range of cells in your spreadsheet that spans 5 rows high and 2 columns wide (e.g., cells A10 through B14).

Enter the [LINEST function](#) formula into the very first cell of the selected range (e.g., cell A10). Ensure you specify the Y-values (known_ys, D2:D8) and the X-values (known_xs, C2:C8), and critically, set the *stats* argument to TRUE to retrieve the full statistics.

The exact formula to input for this specific calculation is:

=LINEST(D2:D8, C2:C8, TRUE, TRUE)

Once the formula is entered and confirmed as an array formula (if required by your Excel version),

the software populates the designated output range with all the regression statistics. The image below displays the result generated by this operation:

	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									

Confirming its consistent location, the **residual sum of squares** is found in the final cell of the second column of the output matrix (cell B14 in this illustration). For this specific simple linear regression model, the residual sum of squares is calculated to be **50.75**. This numerical value encapsulates the model's total aggregate error relative to the actual observations.

Practical Application 2: Calculating RSS for Multiple Linear Regression

The methodology for calculating the **residual sum of squares** for a [multiple linear regression](#) model is remarkably similar to the simple linear case, sharing the same logical framework. The primary difference lies in how the predictor variables (known_xs) are defined within the LINEST function, as a multiple regression utilizes two or more independent variables to predict the single dependent variable.

Let us introduce a second predictor variable, Z, to our existing analysis, creating an expanded 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							

We now aim to model Y as a function of both X and Z. Because we have increased the number of predictor variables to two, the statistical output matrix generated by the LINEST function must expand. Specifically, the output matrix will now be 5 rows high and 3 columns wide ($N = 2$ predictors + 1), requiring a larger selection range (e.g., A10:C14).

The crucial adjustment when applying the LINEST function here is defining the **known_xs** argument. Instead of referencing a single column, we must specify the entire contiguous range that encompasses all predictor columns (both X and Z).

The formula for this multiple linear regression calculation, referencing the expanded predictors C2:D8, would be:

=LINEST(E2:E8, C2:D8, TRUE, TRUE)

Notice that the *known_xs* argument is now C2:D8, correctly encompassing the data for both

predictor variables. The resulting output matrix, spanning 5 rows and 3 columns, is displayed below:

	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											

Despite the expansion of the matrix dimensions, the location of the **residual sum of squares** remains absolutely consistent: it is reliably found in the last cell of the second column of the output area (cell B14 in this example). For this multiple linear regression model incorporating variable Z, the residual sum of squares is determined to be **49.83**.

Interpreting RSS for Model Comparison and Refinement

By comparing the results derived from Example 1 (Simple Linear Regression RSS = 50.75) and Example 2 (Multiple Linear Regression RSS = 49.83), we gain immediate and actionable insight into the impact of adding the second predictor variable (Z). Since the RSS decreased from 50.75 to 49.83, we can definitively conclude that the multiple linear [regression model](#) provides a marginally better fit to the observed data than the simpler model, as it successfully reduced the total unexplained squared error.

It is vital to understand the limitations of RSS as a standalone metric for model comparison. While a lower [residual sum of squares](#) always indicates a closer fit for a given dataset, adding any new predictor variable, even one that is statistically irrelevant, will nearly always reduce the RSS simply due to the nature of the calculation. Therefore, when comparing models with different numbers of

predictors, analysts must also utilize other adjusted metrics, such as the Adjusted R-squared or the F-statistic, to determine if the improvement in fit justifies the increased complexity of the model.

Nonetheless, when comparing two models applied to the exact same dataset and variable combination, a lower **residual sum of squares** is the primary indicator of a superior fit. Mastering the accurate calculation and interpretation of RSS in Excel using the **LINEST function** is an invaluable skill that empowers analysts to swiftly evaluate the performance of their predictive models and make well-informed decisions regarding model refinement and optimal variable selection.