

# How to Calculate RMSE in Excel: A Step-by-Step Guide

Authored by  
**stats writer**

December 29, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Calculate RMSE in Excel: A Step-by-Step Guide*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=109751>

Root Mean Square Error (RMSE) in Excel is a fundamental metric used in statistics and data science to assess the performance of a predictive model. It quantifies the average magnitude of the errors between the predicted values and the observed values. Mathematically, it is defined as the square root of the average of the squared differences between these two sets of data points. The RMSE serves as a robust measure of a model's accuracy, providing a standardized basis for comparing the efficacy of various predictive methods. Achieving a minimal RMSE is the objective when refining a predictive algorithm, as it directly correlates with improved model reliability.

## The Significance of RMSE in Statistical Modeling

In the realm of statistics, regression analysis is a critical methodology employed to determine and analyze the relationship between a dependent variable (often denoted as  $y$ , the response) and one or more independent variables ( $x$ , the predictors). The ultimate output of a regression analysis is a calculated equation, or model, which allows us to forecast the response variable based on the inputs provided by the predictor variables.

Evaluating the goodness-of-fit--how closely the predictions align with reality--is paramount. The Root Mean Square Error (RMSE) provides an indispensable tool for this evaluation. It calculates the standard deviation of the residuals (prediction errors). Because the errors are squared before they are averaged, the RMSE gives disproportionately high weight to large errors, making it sensitive to outliers and a strong indicator of poor model performance.

Essentially, the **RMSE** metric tells us, on average, the distance separating our predicted data points from the actual, observed data points in the dataset. A smaller value indicates that the predictive model is highly accurate and performs well against the given data.

## Understanding the Mathematical Formula for RMSE

The formal mathematical structure used to derive the root mean square error, commonly abbreviated as **RMSE**, precisely captures the process of squaring errors, averaging them, and then reverting the units by taking the square root. This process is necessary to return the error measurement to the original units of the dependent variable, making the metric interpretable.

The formula for calculating the **RMSE** is defined as follows:

$$\mathbf{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Understanding each component of this equation is vital for proper application and interpretation:

$\Sigma$ : This is the summation symbol, indicating that we must calculate the **sum** of all subsequent terms across the entire dataset.

**P<sub>i</sub>**: Represents the **predicted value** generated by the model for the *i*th observation in the dataset.

**O<sub>i</sub>**: Represents the **observed value** (the actual measurement) for the *i*th observation in the dataset.

**n**: This denotes the **sample size**, which is the total number of observations included in the calculation.

### Technical Notes on Terminology:

The root mean square error is versatile and can be utilized for evaluating any type of statistical model that generates predicted values for comparison against real, observed data.

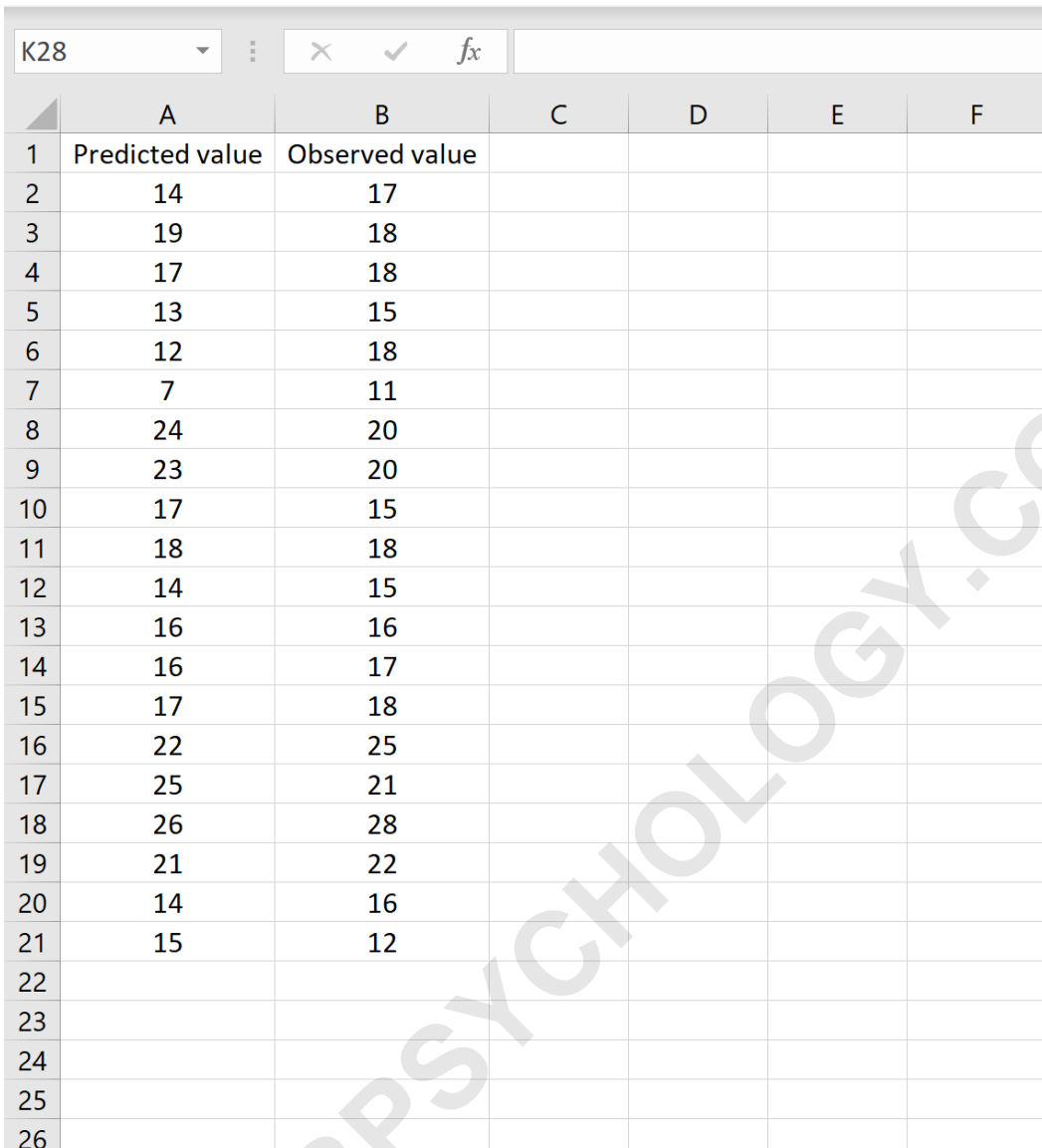
It is important to note that the root mean square error is sometimes interchangeably referred to as the root mean square deviation, often abbreviated as **RMSD**.

## Calculating Root Mean Square Error in Excel

While powerful statistical software packages often have a dedicated function for calculating **RMSE**, Excel does not include a direct, built-in function for this specific metric. However, we can construct the calculation efficiently using a combination of existing functions within a single, powerful **array formula**. This approach allows us to streamline the process significantly, avoiding the need for multiple intermediate calculation columns. We will explore two primary scenarios based on how your data is structured.

### Scenario 1: Using Separate Columns for Predicted and Observed Values

In the most common data arrangement, you will have two distinct columns: one containing the prediction outputs from your regression analysis (Predicted Values) and another detailing the actual observations (Observed Values). The visualization below illustrates this typical dataset setup, where columns A and B hold these respective values for 20 observations:



The screenshot shows an Excel spreadsheet with a grid from column A to F and row 1 to 26. The formula bar at the top shows 'K28' and a function icon. The data is as follows:

	A	B	C	D	E	F
1	Predicted value	Observed value				
2	14	17				
3	19	18				
4	17	18				
5	13	15				
6	12	18				
7	7	11				
8	24	20				
9	23	20				
10	17	15				
11	18	18				
12	14	15				
13	16	16				
14	16	17				
15	17	18				
16	22	25				
17	25	21				
18	26	28				
19	21	22				
20	14	16				
21	15	12				
22						
23						
24						
25						
26						

To calculate the **RMSE** directly in a single cell for this scenario, you must use an array formula. Type the following formula into an empty cell, and crucially, complete the entry by pressing **CTRL+SHIFT+ENTER** (which tells Excel to treat it as an array calculation, inserting curly brackets automatically):

**=SQRT(SUMSQ(A2:A21-B2:B21) / COUNTA(A2:A21))**

Executing this formula correctly yields the RMSE value, demonstrating the model's average error magnitude relative to the data range used.

SUM		=SQRT(SUMSQ(A2:A21-B2:B21)/COUNTA(A2:A21))						
	A	B	C	D	E	F	G	H
1	Predicted value	Observed value						
2	14	17		=SQRT(SUMSQ(A2:A21-B2:B21)/COUNTA(A2:A21))				
3	19	18						
4	17	18						
5	13	15						
6	12	18						
7	7	11						
8	24	20						
9	23	20						
10	17	15						
11	18	18						
12	14	15						
13	16	16						
14	16	17						
15	17	18						
16	22	25						
17	25	21						
18	26	28						
19	21	22						
20	14	16						
21	15	12						
22								
23								
24								
25								
26								

ARABPSYCHOLOGY.COM

	A	B	C	D	E	F	G	H
1	Predicted value	Observed value						
2	14	17		2.6646				
3	19	18						
4	17	18						
5	13	15						
6	12	18						
7	7	11						
8	24	20						
9	23	20						
10	17	15						
11	18	18						
12	14	15						
13	16	16						
14	16	17						
15	17	18						
16	22	25						
17	25	21						
18	26	28						
19	21	22						
20	14	16						
21	15	12						
22								
23								
24								
25								
26								

## Dismantling the Array Formula for Scenario 1

The complex-looking formula achieves the necessary steps of the RMSE calculation--squaring the differences, averaging them, and taking the square root--in a single streamlined operation. Let's dissect how each function contributes to the result:

**=SQRT(SUMSQ(A2:A21-B2:B21) / COUNTA(A2:A21))**

The inner expression, **(A2:A21-B2:B21)**, first calculates the residual (the difference between predicted and observed values) for every corresponding row in the selected ranges.

Next, the **SUMSQ function (SUMSQ())** takes the resulting array of differences, squares each difference, and then calculates the sum of all these squared errors. This fulfills the  $\sum(P_i - O_i)^2$  part of the RMSE formula.

We then divide this sum by the sample size, determined using the **COUNTA()** function, which accurately counts the number of non-empty cells within the range (A2:A21). This step calculates the mean of the squared errors.

Finally, the entire calculation is enclosed within the **SQRT()** function, which takes the square root of the mean squared error, yielding the final **RMSE** value.

## Scenario 2: Utilizing Pre-calculated Difference Columns

In some analyses, the residuals (the differences between predicted and observed values) may have already been calculated and placed into a dedicated column. In this scenario, you only have a single column representing the errors, simplifying the array formula slightly.

The example below illustrates this situation: Predicted values are in column A, Observed values are in column B, and the residual (A minus B) has been pre-calculated and placed into column D:

	A	B	C	D	E	F
1	Predicted value	Observed value		Predicted - Observed		
2	14	17		-3		
3	19	18		1		
4	17	18		-1		
5	13	15		-2		
6	12	18		-6		
7	7	11		-4		
8	24	20		4		
9	23	20		3		
10	17	15		2		
11	18	18		0		
12	14	15		-1		
13	16	16		0		
14	16	17		-1		
15	17	18		-1		
16	22	25		-3		
17	25	21		4		
18	26	28		-2		
19	21	22		-1		
20	14	16		-2		
21	15	12		3		
22						
23						
24						
25						

If your data is structured this way, the array formula focuses solely on column D. Enter the following formula into an empty cell and remember to press **CTRL+SHIFT+ENTER** to execute the array calculation:

**=SQRT(SUMSQ(D2:D21) / COUNTA(D2:D21))**

	A	B	C	D	E	F	G	H
1	Predicted value	Observed value		Predicted - Observed				
2	14	17		-3	=SQRT(SUMSQ(D2:D21) / COUNTA(D2:D21))			
3	19	18		1	SQRT(number)			
4	17	18		-1				
5	13	15		-2				
6	12	18		-6				
7	7	11		-4				
8	24	20		4				
9	23	20		3				
10	17	15		2				
11	18	18		0				
12	14	15		-1				
13	16	16		0				
14	16	17		-1				
15	17	18		-1				
16	22	25		-3				
17	25	21		4				
18	26	28		-2				
19	21	22		-1				
20	14	16		-2				
21	15	12		3				
22								
23								
24								
25								

This calculation yields the root mean square error of **2.6646**. Notably, this result precisely matches the outcome from Scenario 1, confirming that both methodological approaches for calculating **RMSE** are mathematically equivalent and valid.

	A	B	C	D	E	F
1	Predicted value	Observed value		Predicted - Observed		
2	14	17		-3	2.664583	
3	19	18		1		
4	17	18		-1		
5	13	15		-2		
6	12	18		-6		
7	7	11		-4		
8	24	20		4		
9	23	20		3		
10	17	15		2		
11	18	18		0		
12	14	15		-1		
13	16	16		0		
14	16	17		-1		
15	17	18		-1		
16	22	25		-3		
17	25	21		4		
18	26	28		-2		
19	21	22		-1		
20	14	16		-2		
21	15	12		3		
22						
23						
24						
25						

## Deconstructing the Formula for Pre-calculated Differences

While conceptually simpler, the structure of the formula for Scenario 2 still relies on the essential combination of **SUMSQ()** and **COUNTA()** functions:

**=SQRT(SUMSQ(D2:D21) / COUNTA(D2:D21))**

Since column D already contains the residuals, the **SUMSQ function (SUMSQ())** is applied directly to the range D2:D21. It squares each existing difference and sums these squared values, fulfilling the numerator of the RMSE equation.

The result is then divided by the number of observations, which is obtained using the **COUNTA()** function, counting the data points in column D.

The overarching **SQRT()** function ensures that the final error metric is returned to the original units of measurement.

## Interpreting the Resulting RMSE Value

As established, **RMSE** is a highly effective tool for gauging the fit of a predictive model, regardless of whether it originates from linear regression analysis or another predictive methodology. Interpreting the magnitude of the resulting value is crucial for actionable analysis.

A high **RMSE** value signifies a large average difference between the predicted and observed data points. This result indicates that the statistical model is poorly calibrated or suffers from significant error, meaning it does not fit the underlying data distribution well. Conversely, a minimal **RMSE** value demonstrates strong correlation and high fidelity between predictions and reality, suggesting the model is robust and accurate.

One of the most valuable uses of the RMSE is its use in **model comparison**. When evaluating two or more models built on the same dataset, the model that yields the lowest RMSE is statistically considered the superior predictor. This comparison allows analysts to systematically select the most appropriate algorithm for forecasting tasks, ensuring optimal performance and reliability.

***For more detailed tutorials in Excel, be sure to check out our resource page, which lists every Excel tutorial available on this platform.***