

# Learning Root Mean Square Error (RMSE) and Calculation Guide in Excel

Authored by  
**Mohammed loot**

November 8, 2025

## RECOMMENDED CITATION

Mohammed loot (2025). *Learning Root Mean Square Error (RMSE) and Calculation Guide in Excel*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=13896>

In the realm of statistics and data science, [regression analysis](#) stands as a foundational technique. It is critically important for exploring and quantifying the relationship between a set of predictor variables (independent variables, commonly represented as  $\mathbf{x}$ ) and a response variable (the dependent variable,  $\mathbf{y}$ ). Through this robust analytical process, researchers and analysts are able to construct powerful [predictive models](#) capable of forecasting potential outcomes based on a given set of input data.

Successful execution of a regression analysis yields a mathematical structure--often represented as a line or curve--that optimally describes the relationship within the observed data points. This structure is then leveraged to generate a **predicted value** for the response variable given any input from the predictor variables. Crucially, since real-world data is seldom perfect and always subject to noise, achieving absolute prediction accuracy is impossible. Therefore, rigorously assessing the accuracy and magnitude of prediction errors is an essential step for validating the practical utility and robustness of the resulting model.

The most essential and universally accepted metric for quantifying model fit is the [root mean square error \(RMSE\)](#). The **RMSE** provides a singular, highly interpretable measure of the average magnitude of prediction errors. Specifically, it represents the standard deviation of the [residuals](#) (the errors) and indicates the typical distance separating the model's **predicted values** from the actual, corresponding [observed values](#) in the dataset. Fundamentally, a lower **RMSE** value is the direct indicator of a more precise and superior-fitting predictive model.

## The Mathematical Definition of RMSE

Calculating the **root mean square error** is a robust, four-step process. It begins with determining the [residuals](#) (the difference between predicted and observed values), followed by squaring those residuals, computing the mean of the squared residuals, and concluding with taking the square root of that mean. This ensures that the metric is in the same units as the response variable. This rigorous mathematical procedure is summarized by the standard formula below:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2}$$

For accurate implementation and meaningful interpretation of the **RMSE**, a clear understanding of each component within the formula is paramount. These elements define how the error is aggregated and normalized:

$\Sigma$  represents the operation of **summation**, instructing us to aggregate all calculated squared errors  $(P_i - O_i)^2$  across the entire dataset.

$P_i$  is the **predicted value** output by the statistical model for the  $i$ th data point.

$O_i$  is the corresponding [observed value](#) (the actual measurement or ground truth) for the  $i$ th data point.

$n$  denotes the **sample size**--the total number of observations included in the calculation.

### Technical Notes on RMSE:

The utility of the **root mean square error** is not limited to traditional regression; it is versatile enough to be applied to any **statistical model**, including advanced time series forecasting or complex machine learning algorithms, as long as a comparison between **predicted values** and **observed values** can be made.

While **RMSE** is the standard abbreviation, this metric is frequently encountered in academic and technical texts as the Root Mean Square Deviation (RMSD). Both terms refer to the identical statistical measure of error magnitude.

Although the theoretical basis for **RMSE** is sound, performing this calculation manually, especially when dealing with large datasets, is time-consuming and error-prone. To achieve efficiency and accuracy, we turn to powerful spreadsheet software. The following sections provide a practical, detailed, step-by-step methodology for calculating the **root mean square error** directly within Microsoft Excel, covering common data structuring scenarios.

## How to Calculate Root Mean Square Error in Excel

It is important to recognize that Microsoft Excel does not include a single, native function explicitly named "RMSE." Despite this, analysts can readily construct the necessary calculation by combining several powerful, built-in array functions. We will detail two common data arrangement scenarios and demonstrate how to derive the correct **RMSE** result for both. Crucially, these methods rely on deploying an **array formula**, which must always be finalized by pressing the **CTRL+SHIFT+ENTER** sequence instead of just Enter.

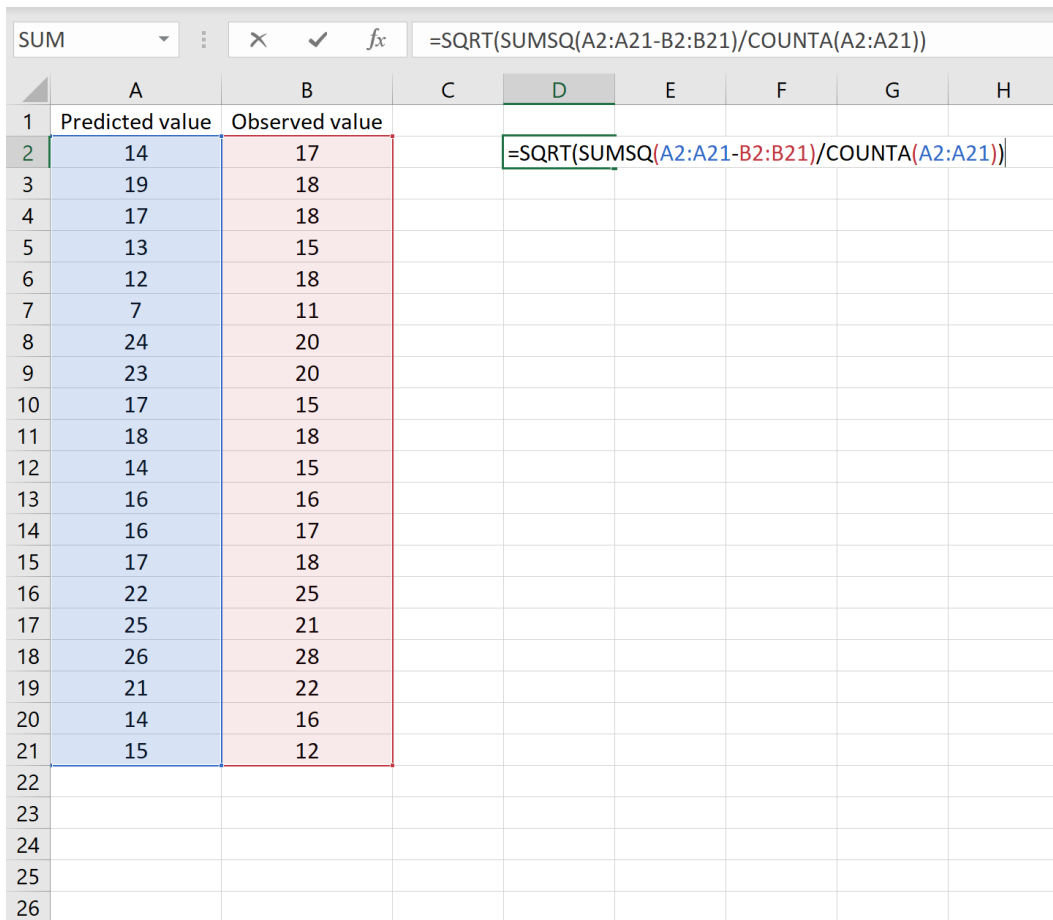
### Scenario 1: Separate Columns for Observed and Predicted Values

The typical data arrangement involves two adjacent columns: one dedicated to the model's **predicted values** and the other holding the actual **observed values**. This structure is ideal because it permits the calculation of the **residuals** (the raw differences) and their subsequent summation within one highly efficient array formula. The illustration below demonstrates this standard configuration, where Predicted Values occupy Column A and Observed Values are in Column B:

	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 accurately compute the **RMSE** for this layout, an array formula is mandatory. Input the formula provided below into a suitable blank cell. The crucial step is finalizing the entry: instead of simply pressing Enter, you must simultaneously press **CTRL+SHIFT+ENTER**. This specific command is essential because it signals Excel to treat the ranges (A2:A21 and B2:B21) as arrays, enabling simultaneous subtraction, squaring, and summing of all data points in one operation.

**=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								

Upon correct execution as an array formula (indicated by curly braces appearing around the formula in the formula bar, `{=SQRT(SUMSQ(...))}`), the final RMSE value will be instantly displayed. This outcome showcases the efficiency gained by integrating Excel's specialized mathematical and array functions.

	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								

While the combined formula might initially seem complicated, a closer examination reveals that its logical structure mirrors the exact mathematical definition of **RMSE**. The calculation can be broken down into three core functional components, which address the necessary steps of squaring errors, calculating the mean, and taking the root:

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

The term `(A2:A21-B2:B21)` generates an array of differences, or **residuals**, for every row. The inner **SUMSQ()** function then performs two critical actions: it squares each residual and computes the sum of these squared differences, satisfying the numerator  $(\sum(P_i - O_i)^2)$  of the RMSE equation. This sum is then divided by the total count of observations, which is determined by the **COUNTA()** function. The **COUNTA()** function reliably counts all non-blank cells in the specified range (A2:A21), thereby calculating the effective **sample size** (n) used for normalization.

Finally, the outer **SQRT()** function executes the last step: taking the square root of the entire result. This converts the mean squared error back into the original units of measurement, providing the final, interpretable **RMSE** value.

## Scenario 2: Calculating RMSE from Pre-Calculated Residuals

A slightly different scenario arises when your analytical workflow has already produced a column containing the raw differences, or **residuals**, between the **predicted values** and the **observed values**. This pre-calculation streamlines the Excel formula significantly because the internal array subtraction step ( $P_i - O_i$ ) is rendered unnecessary.

The following image demonstrates this alternative setup. Columns A and B still contain the Predicted and Observed data, respectively, but the pre-calculated residuals (the difference B-A) are now explicitly listed in 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						

When utilizing this structure, the formula simplifies to reference only the column containing the residuals (D2:D21). Enter the following formula into your chosen calculation cell, making sure to finalize the input by pressing **CTRL+SHIFT+ENTER** to ensure it functions correctly as an 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								

Upon successful execution, the calculation confirms a **root mean square error** is exactly **2.6646**. This result is identical to the one derived in Scenario 1, proving that both methodologies--calculating the differences implicitly within the array formula or explicitly referencing a column of pre-calculated residuals--are mathematically equivalent and reliable for determining model error.

	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						

The underlying structure of the formula for Scenario 2 remains consistent with the mathematical definition, differing only in its direct reference to the residual column (D):

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

As the differences are already established in Column D, the numerator calculation is streamlined. The **SUMSQ()** function is applied directly to the range D2:D21, instantaneously squaring each residual and summing these squared differences.

This summed result is then normalized by dividing it by the total count of data points (n), which is robustly calculated using the **COUNTA()** function across the residual range.

The final component, the **SQRT()** function, finishes the process by taking the square root, transforming the mean squared error back into the original scale of the **observed values**.

## How to Interpret Root Mean Square Error

The **RMSE** is an indispensable diagnostic measure, offering concrete proof of how effectively a **statistical model** captures the inherent patterns within a dataset. Its interpretation is straightforward and highly practical because the resulting metric is always expressed in the exact same units as the original response variable (or the **observed values**). This inherent scaling allows for easy comparison with the input data.

The fundamental principle guiding **RMSE** interpretation is that the size of the value directly corresponds to the average size of the prediction error. A substantially high **RMSE** indicates a large average discrepancy between the model's forecasts and the actual outcomes, suggesting a poor fit to the data. Conversely, a low **RMSE** value is highly desirable, as it signifies that the model's **predicted values** closely align with the true data points, translating to high predictive accuracy and overall model robustness.

Perhaps the most powerful application of **RMSE** lies in its utility for **model comparison**. When analysts need to choose among two or more distinct **statistical models** attempting to solve the same predictive challenge, comparing their respective RMSE values provides a definitive, objective metric for identifying superior performance. The rule is simple: the model that produces the lowest **RMSE** is quantitatively the best choice, as it minimizes the average prediction error and, consequently, is the most reliable forecasting instrument available.

***For more detailed tutorials in Excel and statistical analysis, be sure to check out our comprehensive library, which lists every Excel tutorial on Statology.***