

# Learning Multiple Linear Regression with Excel's LINEST Function

Authored by  
**Mohammed Iooti**

October 27, 2025

## RECOMMENDED CITATION

Mohammed Iooti (2025). *Learning Multiple Linear Regression with Excel's LINEST Function*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=4470>

The [LINEST function](#) in Microsoft Excel stands out as an exceptionally powerful utility for rigorous statistical analysis. Specifically, it is designed to facilitate the fitting of a [multiple linear regression](#) model, enabling analysts to quantify the relationship between a single outcome (dependent) variable and two or more influencing (independent) variables. This capability moves beyond simple data visualization, offering essential quantitative insights necessary for accurate predictive modeling, forecasting, and deep data interpretation across various professional disciplines.

Mastering the effective utilization of [LINEST](#) is a significant step toward elevating your data analysis proficiency within the Excel environment. Unlike basic trendline analysis, which handles only one predictor, LINEST provides a robust, multivariate methodology to precisely quantify the simultaneous impact of multiple factors on a specific outcome. This comprehensive guide will systematically detail the entire process, starting from the fundamental function structure, navigating through data preparation, model implementation, and culminating in the detailed interpretation of both basic and advanced [regression statistics](#), ensuring you can confidently apply multiple linear regression in your complex analytical projects.

## Understanding the LINEST Function Syntax

To ensure accurate computation and implementation, the [LINEST function](#) relies on a precise [syntax](#). A clear understanding of this structure is paramount, as it dictates how Excel receives your raw data inputs and specifies the exact statistical output required for your desired regression analysis. The function requires users to meticulously define which data range represents the variable being modeled and which ranges represent the variables used for prediction.

The fundamental structure of the function, including both mandatory and optional parameters, is provided below. This formula is designed to be executed as an array function, meaning it is capable of returning multiple results (the [coefficients](#) and optional statistics) into a designated range of cells.

**=LINEST(known\_y's, , , )**

Each argument within this structure is critical for defining the specific linear relationship that Excel will attempt to compute and fit to your data. Understanding the role of each component ensures that the resulting [regression model](#) accurately reflects the hypotheses being tested. We must now explore the detailed purpose of each argument to ensure comprehensive implementation knowledge.

**known\_y's:** This is a mandatory argument representing the array or cell range containing the known dependent y-values. These are the observed values of the [response variable](#)--the variable whose variation the model attempts to predict or explain based on the independent variables.

**known\_x's:** This optional argument specifies the array or range of known independent x-values. These ranges constitute the [predictor variables](#), which are hypothesized to exert influence over the y-values. If this argument is omitted, Excel automatically defaults to assuming a simple sequence of x-values: {1, 2, 3, ...} corresponding to the number of y-observations provided.

**const:** This is an optional logical argument determining whether the [intercept](#) 'b' (the constant term in the regression equation) should be included in the calculation. If set to **TRUE** (or omitted), the intercept is calculated normally. If set to **FALSE**, the regression line is forced through the origin (0, 0) or, more accurately, the intercept 'b' is set to zero, meaning the model assumes the response is zero when all predictor variables are zero.

**stats:** This is another optional logical argument controlling the comprehensiveness of the output. Setting this to **TRUE** instructs the function to return a full range of additional [regression statistics](#), offering a more detailed assessment of the model's reliability and the precision of the calculated [coefficients](#). If set to **FALSE** or omitted, the function returns only the slope coefficients and the intercept value.

A thorough grasp of these arguments is the foundational prerequisite for successfully applying the [LINEST function](#), enabling analysts to correctly structure their model and extract the precise insights required. The next section transitions to a practical, step-by-step example, demonstrating how to prepare data and input these arguments effectively into a real-world scenario.

## Preparing Your Data for Analysis

Prior to initiating any sophisticated statistical computation, the correct organization of your data within the Excel spreadsheet is absolutely fundamental. A meticulously structured [dataset](#) is essential for ensuring that the [LINEST function](#) can unambiguously map the ranges to the correct [response variable](#) (Y) and the corresponding [predictor variables](#) (X's), thereby guaranteeing the validity and accuracy of the resulting regression analysis.

For our hands-on practical demonstration of [multiple linear regression](#), we will begin by populating a sample [dataset](#) into an Excel worksheet. This sample dataset is formulated to represent a typical business or scientific scenario, featuring three distinct [predictor variables](#) (labeled x1, x2, and x3) and a single designated [response variable](#) (y), which we aim to predict and model the relationship for.

To accurately replicate the subsequent steps and obtain the demonstrated results, please ensure your data is entered into the specified columns and rows exactly as depicted in the following visual aid. Maintaining the precise layout--with predictors grouped together and the response variable segregated--is crucial for defining the formula ranges correctly in the next stage.

	A	B	C	D	E	F
1	<b>x1</b>	<b>x2</b>	<b>x3</b>	<b>y</b>		
2	1	8	7	8		
3	2	7	5	12		
4	2	7	5	14		
5	2	7	4	15		
6	3	6	6	15		
7	4	6	5	14		
8	6	4	6	19		
9	5	5	8	29		
10	6	6	10	24		
11	5	5	14	26		
12	7	4	7	22		
13	9	2	6	30		
14	8	1	4	34		
15						
16						
17						
18						
19						
20						
21						

The integrity of the statistical outcome hinges on careful and accurate data entry; it is a fundamental prerequisite for any reliable analysis. Once you have confirmed that your data is correctly populated and structured within the Excel sheet, we are prepared to move forward to the core task: implementing the [LINEST function](#) and generating the initial regression results.

## Implementing Multiple Linear Regression with LINEST

With our input data now meticulously organized, the next critical phase involves constructing and executing our [multiple linear regression](#) model using the [LINEST function](#). Our primary analytical objective here is to derive an equation that can accurately predict the values of the [response variable y](#) by assessing the combined, weighted influence of the three [predictor variables](#): **x1**, **x2**, and **x3**.

To accomplish this, we will input a specific formula into an empty cell on our worksheet. Given the layout established in the previous section--where the 'y' values reside in column D and the 'x' values collectively span columns A through C--the formula must reference these ranges appropriately. It is vital to remember that [LINEST](#) is an array function; consequently, it is designed to spill its multiple output values across a range of cells. In legacy versions of Excel (pre-2019), the

formula must be confirmed by pressing **Ctrl+Shift+Enter**. In modern versions (Excel 365 or 2019+), simply entering the formula in the top-left cell of the desired output area is sufficient.

Place your cursor in any empty cell where you wish the output to begin and enter the following formula:

**=LINEST(D2:D14, A2:C14)**

In the syntax above, `D2:D14` corresponds directly to our **known\_y's** (the response variable 'y'), and `A2:C14` designates our **known\_x's** (the [predictor variables](#)  $x_1$ ,  $x_2$ , and  $x_3$ ). Since we have deliberately omitted the optional `const` and `stats` arguments, Excel defaults to assuming `TRUE` for `const` (calculating the [intercept](#) normally) and `FALSE` for `stats`, resulting in a streamlined output that includes only the calculated [coefficients](#) and the [intercept](#) value.

The screenshot below visually confirms the application of this formula in the Excel interface and displays the immediate, compact array of values returned by the function. Notice that the output is automatically spread horizontally across four adjacent cells, corresponding to the three predictors plus the intercept.

	A	B	C	D	E	F	G	H	I
1	x1	x2	x3	y		0.849687	-3.00393	0.34271	28.5986
2	1	8	7	8					
3	2	7	5	12					
4	2	7	5	14					
5	2	7	4	15					
6	3	6	6	15					
7	4	6	5	14					
8	6	4	6	19					
9	5	5	8	29					
10	6	6	10	24					
11	5	5	14	26					
12	7	4	7	22					
13	9	2	6	30					
14	8	1	4	34					
15									
16									
17									
18									
19									
20									
21									
22									
23									

The values generated by [LINEST](#) are the estimated [coefficients](#) that define the linear equation. Crucially, they are ordered from right to left: the coefficient for the last x-variable (x3), followed by x2, x1, and finally, the [intercept](#) term. This specific right-to-left order is standard and must be remembered when interpreting the results of the LINEST function.

## Interpreting the Regression Output

The array of numbers returned by the [LINEST function](#) represents the essential [coefficients](#) that mathematically define our [multiple linear regression](#) model. The real value of this analysis lies in translating these numerical estimates into meaningful, actionable conclusions about the relationships between our variables. Each coefficient quantifies the expected change in the response variable associated with a one-unit change in its corresponding predictor, assuming all other predictors remain constant (the *ceteris paribus* assumption).

By referencing the output obtained in the previous implementation step, we can isolate and interpret the specific [coefficients](#) derived for our model. This interpretation allows us to precisely

articulate the fitted linear relationship, detailing the magnitude and direction (positive or negative) of the influence each [predictor variable](#) has on the response.

The [coefficient](#) for the [intercept](#) (b) is **28.5986**. This value represents the baseline--the predicted mean value of 'y' when all [predictor variables](#) (x1, x2, x3) are hypothetically held at zero.

The [coefficient](#) for x1 is **0.34271**. This positive value indicates that, holding x2 and x3 constant, for every one-unit increase observed in x1, the [response variable](#) y is expected to increase by approximately 0.34271 units.

The [coefficient](#) for x2 is **-3.00393**. This negative sign suggests an inverse relationship: for every one-unit increase in x2, while x1 and x3 are held constant, y is expected to decrease by approximately 3.00393 units.

The [coefficient](#) for x3 is **0.849687**. This positive value means that a one-unit increase in x3 is associated with an expected increase of about 0.849687 units in y, assuming x1 and x2 are unchanged.

By integrating these calculated [coefficients](#), we can formally construct the complete mathematical representation of the fitted [regression equation](#). This equation is the heart of the predictive model:

$$y = 28.5986 + 0.34271(x1) - 3.00393(x2) + 0.849687(x3)$$

This formula is now a powerful tool, allowing you to substitute any combination of observed or hypothesized values for x1, x2, and x3 (within the realistic range of your training data) to predict the corresponding value of 'y'. It provides a definitive quantitative description of the linear relationship established between the [response variable](#) and the set of [predictor variables](#).

## Unlocking Advanced Regression Statistics

While the basic output of [LINEST](#) successfully provides the model's core [coefficients](#), this limited view does not offer sufficient information to rigorously validate the model's quality or the reliability of individual coefficient estimates. By setting the optional **stats** argument to **TRUE**, we instruct Excel to unlock and display a comprehensive wealth of additional [regression statistics](#). These statistics are fundamentally important for accurately evaluating the overall goodness of fit, assessing statistical significance, and determining the precision of our [multiple linear regression](#) model.

To generate this comprehensive statistical report, you must modify the initial [LINEST function](#) call to explicitly include **TRUE** for both the `const` and `stats` arguments. This modification signals to Excel that the full, detailed output matrix is required.

**=LINEST(D2:D14, A2:C14, TRUE, TRUE)**

It is crucial that this formula be entered as an array function spanning a sufficiently large range of cells to accommodate the full output matrix. For a multiple linear regression with three predictor variables, the advanced output requires 5 rows and 4 columns ( $k+1$  columns for coefficients/intercept and 5 rows for the statistics). The following screenshot visually represents the expanded output matrix returned after successfully executing the formula:

	A	B	C	D	E	F	G	H	I
1	x1	x2	x3	y		0.849687	-3.00393	0.34271	28.5986
2		1	8	7	8	0.453295	1.626423	1.327566	13.20088
3		2	7	5	12	0.838007	3.707539	#N/A	#N/A
4		2	7	5	14	15.51935	9	#N/A	#N/A
5		2	7	4	15	639.9797	123.7126	#N/A	#N/A
6		3	6	6	15				
7		4	6	5	14				
8		6	4	6	19				
9		5	5	8	29				
10		6	6	10	24				
11		5	5	14	26				
12		7	4	7	22				
13		9	2	6	30				
14		8	1	4	34				
15									
16									
17									
18									
19									
20									

While the fundamental fitted [regression equation](#)--the top row of the output--remains unchanged, the subsequent rows provide vital context regarding the model's performance and the precision of the coefficient estimates. Understanding the meaning of each statistic is paramount to confirming the model's reliability.

Here is a detailed breakdown and interpretation of the key additional [regression statistics](#) provided in the expanded output matrix:

The [standard error](#) (SE) for  $x_3$  (**0.453295**),  $x_2$  (**1.626423**),  $x_1$  (**1.327566**), and the [intercept](#) (**13.20088**) is listed in the second row. The SE measures the average expected variation of the calculated coefficient from the true population value across different samples. A lower standard error implies a more precise and reliable estimate for that specific coefficient.

The [R<sup>2</sup>](#) value, located in the third row, first column, is **.838007**. Also known as the [coefficient of](#)

**determination**, this crucial metric quantifies the proportion of the total **variance** observed in the **response variable** 'y' that is successfully explained by the collective set of **predictor variables** (x1, x2, x3) included in the model.

The **residual standard error** for y, found in the third row, second column, is **3.707539**. This value is also referred to as the standard error of the regression, and it serves as a measure of the average distance that the actual observed y-values fall from the estimated regression line. It represents the typical magnitude of the prediction errors (residuals).

The overall **F-statistic** (fourth row, first column) is **15.51925**. This statistic is fundamental for testing the overall statistical significance of the entire **regression model**. A high F-statistic strongly suggests that the model, as a whole, is significantly better at predicting the response variable than a null model that uses only the mean of Y.

The **degrees of freedom** (df), located in the fourth row, second column, is **9**. In this context, it refers to the residual degrees of freedom, calculated as  $(n - k - 1)$ , where 'n' is the number of observations and 'k' is the number of **predictor variables**. This value is used in statistical testing.

The **regression sum of squares** (RSS, fifth row, first column) is **639.9797**. This measures the portion of the total variation in the **response variable** that is successfully accounted for and explained by the linear relationship captured by the **regression model**.

The **residual sum of squares** (fifth row, second column) is **123.7126**. This quantifies the remaining variation in the **response variable** that the **regression model** fails to explain, representing the sum of the squared prediction errors.

Among this array of detailed statistics, the **R2 value** is typically the most frequently examined metric for rapidly assessing the overall goodness of fit of the predictive model. It provides a simple, scaled measure (ranging from 0 to 1) of how effectively the **regression model** explains the observed **variance** in the **response variable**.

Since the **R2** for this specific model is calculated as **.838**, we can confidently assert that approximately 83.8% of the total **variance** in the **response variable** 'y' is explained by the combination of the **predictor variables** (x1, x2, and x3). This result suggests a very strong fit, confirming that the selected predictors are highly effective and collectively powerful in forecasting the outcome variable 'y' within the context of the observed data.

## Conclusion and Further Exploration

The **LINEST function** within Excel provides data professionals and analysts with a highly robust and remarkably accessible method for executing complex **multiple linear regression** analysis. By diligently following the procedural steps detailed in this guide--from careful data preparation to formula implementation--users can not only derive the fundamental **coefficients** required for their predictive models but also gain a profound, quantitative understanding of the model's statistical validity and overall explanatory power by utilizing the optional statistical output.

Mastering the intricacies of this function is transformative, enabling you to transition beyond simple data summarization and aggregation into the realm of sophisticated predictive analytics and forecasting. This advanced capability facilitates significantly more informed, data-driven decision-making across diverse professional fields, including finance, engineering, and social science research. It is important to bear in mind that, despite the power of [LINEST](#), any sound statistical analysis must be coupled with a critical, nuanced understanding of the underlying statistical assumptions, such as linearity and homoscedasticity, and the specific context of the [dataset](#) being analyzed.

We strongly encourage practitioners to continue experimenting with various [datasets](#) and actively exploring the distinct impacts of manipulating the optional `const` and `stats` arguments. Such hands-on exploration will further solidify your technical comprehension of the [LINEST function](#) and fully unlock its powerful potential for advanced modeling within the Excel environment.

## **Additional Resources for Excel Analysis**

For those seeking to further enhance their statistical and analytical expertise within Excel, the following tutorials offer guidance on other common operations and related functions, providing pathways to deepen your overall data science toolkit: